

Germans engage with physics

In an attempt to reverse a declining interest in scientific careers and to remind the public that their work is interesting, German researchers have succeeded in enhancing the accessibility of physics. Others should benefit from their example.

The principle that new science should be understandable to the general public was embedded in the philosophy of the Age of Enlightenment. In 1738, Voltaire published his *Elément de la philosophie de Newton: mis à la portée de tout la monde*, but *tout la monde* was in practice often restricted to that section of the European public deemed to be most in need of simple explanation: women. Francesco Algarotti's *Il newtonismo per le dame* (1739) was one of many pamphlets produced for 'ladies of leisure'.

Nowadays, science's 'public' has many components, including influential people who want science to be more accountable than it has been in the past, who resent being patronized by scientists and who reject the notion that understanding leads inevitably to support. Making science accessible, rather than merely understandable, is now as much a political duty as a philosophy. Germany's science community, a late starter in taking this seriously, has to reach out to a particularly distrustful public. The signs are that it is rising to the challenge in an exemplary fashion.

Two years ago, research organizations, prodded into action by a new generation of dynamic science policy leaders, started an initiative dubbed "the Public Understanding of Science and Humanities". The mistake of using that English title has since been corrected to *Wissenschaft im Dialog* — a more appropriate choice of words as well as language. The organizers stress that dialogue rather than acceptance is the goal, although they clearly believe that greater acceptance will follow.

Wissenschaft im Dialog, led by Joachim Treusch of Forschungszentrum Jülich, is supported practically and financially by German research organizations and foundations, as well as the federal research ministry. It was right to hire a public relations consultancy to help run and assess the programmes. The company has not only implemented the programmes designed by the scientists but has also suggested additional elements that the scientists would have been too inexperienced to think of themselves. Half-way through the first major programme of events, the "Year of Physics", the signs are that the formula is working.

The Year of Physics was first conceived by the German Physical Society (DPG), which had been worried by an alarming drop in the number of physics students and the looming gap between the number of trained physicists and employers' needs. The DPG remains the organizer, but the initiative now lies under the umbrella of *Wissenschaft im Dialog*, and has spawned the Year of Life Sciences (2001) and the Year of Geosciences (2002).

The events of the Year of Physics have been imaginative and wide-ranging. The year has been divided into five parts, each dedicated to a different physics discipline. Particle physicists are still glowing with pride at the record attendance they attracted in April to *Reise zum Urknall*, a series of public lectures accompanying exhibitions, one provided by CERN, taking visitors back to the Big Bang. The week-long event attracted some 15,500 visitors to the Urania, Berlin's nineteenth-century public lecture hall.

In early summer, the exhibition *Gebändigtes Licht*, celebrating quantum optics, saw crowds squashing into a marquee in the main square in Bonn well into the small hours on three consecutive nights. Physicists took their experiments into Bonn's department stores and banks, making active contact, they believe, with tens of thousands of 'ordinary' — and no doubt rather surprised — folk.

Other events during the year include artists-meet-science as well as hands-on 'physics is fun' experiments for children, and a season of Hollywood films with scientific themes. The project ends in Berlin in December, coinciding with celebrations of the 100th anniversary of Max Planck's quantum theory.

The Year of Physics has attracted wide media coverage. Moreover, hundreds of thousands of the public, which scientists perceive to be increasingly receptive to science, will have encountered it directly. The significant progress already made is a tribute both to researchers' dedication to the task and to the professionalism being adopted. That combination will be even more necessary, but stands to gain accordingly, when the Year of Life Sciences reaches out to a public notable for its genome phobia (see page 821).

Stem cells democracy

Procrastination indicates an inappropriate lack of faith in political processes.

One of the most surprising aspects of last week's decision by the British government to recommend that research be allowed on embryonic stem cells (see page 815) is not the decision itself, but the length of time it took to emerge. After all, the principle that human embryos up to 14 days old can be used in research was approved by parliament ten years ago, after lengthy public debate, when it passed the Human Fertilization and Embryology Act. A new parliamentary decision is only needed now because, in explicitly listing the types of research covered by the act, work on stem cells was not included — for the simple reason that such research was unknown at the time.

The delay is being widely interpreted as reflecting the government's nervousness about the likely public reaction to its proposal. Ever since the political fiasco over the safety of genetically modified crops, there has been an air of uncertainty hanging over how to

handle sensitive decisions involving a potential clash between scientific advances — particularly those related to health and reproduction — and public sentiment. The case of BSE showed the world how this should not be done.

In contrast, the British debate on research using embryos, which has involved a lengthy dialogue between scientific, legal, religious and ethical experts, open participation by members of the public, and will culminate in a free vote in parliament, is a case study in how to do things correctly. So, although the government's caution is commendable, the appearance of hesitation is less so. Researchers have confidence in the value of the science — even if only as a necessary step towards what is being described as the 'Holy Grail' of reprogramming adult stem cells. Politicians should have equal confidence in the robustness of the democratic process.





Green giant

The world's largest single-dish telescope is unveiled
p816



Dioxin danger

Plans for joint US–Vietnamese research projects
p818



Out in the cold

Doctors are unhappy with Iceland's new data collection rules
p819



Racing ahead

Local research in Siena boosted by world's first bank
p821

UK government backs change in law over stem-cell research...

David Dickson, London
& Paul Smaglik, Washington

The British government last week urged the country's research councils to draw up a programme of research into the potential therapeutic benefits to be derived from using human embryonic stem cells to replace diseased or damaged tissue.

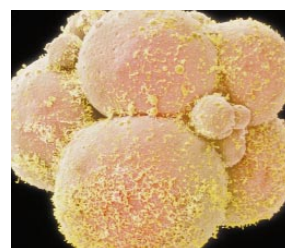
At the same time, it has recommended that parliament modify an act passed ten years ago, which already permits research on embryos up to 14 days old, to allow embryo stem-cell research to take place in the United Kingdom.

But the government will allow members of parliament a free vote on the issue — a relatively unusual move in a political system where most votes take place along political lines. And, in a bid to draw a firm line between what is and what is not permitted, it has also promised new legislation explicitly outlawing the reproductive cloning of human beings.

The move comes as the US National Institutes of Health prepares to issue guidelines — expected by some in the next few days — on the conditions under which it is prepared to finance human embryonic stem-cell research. But the longer term future of federal funding of such research hinges on the outcome of the US presidential campaign.

The British moves have brought protests from the country's vocal anti-abortion lobby which, like its counterpart in the United States, is arguing that the same goals could eventually be achieved through research using adult stem cells, thus, they claim, avoiding the need to use embryos.

Elsewhere in Europe, the reaction was equally mixed. Although the Vatican, perhaps predictably, issued a highly critical statement on the moral implications, others offered a milder rebuke to the British government for seeking to press ahead unilaterally with a 'functionalist' approach to embryonic



Huge potential: embryonic stem cells could revolutionize medicine.

stem-cell research, rather than trying to achieve a Europe-wide consensus first.

In general, the government's move has met with approval both in Britain — where only one newspaper, the conservative *Daily Telegraph*, published a critical editorial — and abroad, where it has been widely welcomed as a pragmatic effort to balance ethical concerns for the treatment of embryos against the medical benefits likely to emerge from the research.

The decision came in response to proposals contained in a report drawn up over the past year by a panel headed by chief medical officer Liam Donaldson. This followed earlier recommendations to allow embryonic stem-cell research, leading eventually to 'therapeutic cloning', from both the Human Fertilisation and Embryology Authority and the Human Genetics Advisory Commission, subsequently endorsed by the Nuffield Council on Bioethics (see *Nature* **404**, 697; 2000).

In announcing that the government had accepted his panel's proposals, Donaldson noted that although the possibility of 're-programming' adult cells was the 'Holy Grail' of stem-cell research, at present the use of either adult or fetal cells "are thought to be of more limited potential". He added: "It is partly a question of exploiting embryonic stem cells and seeing their potential, but also about resolving some of the theoretical, technical and safety questions."

It is thought that between five and six research groups in Britain are contemplating applying for funding for such research. The Donaldson report also calls for consideration of establishing collections of stem cells for research use, and the government has called on the research councils to do this. ■

♦ <http://www.doh.gov.uk/cegc>

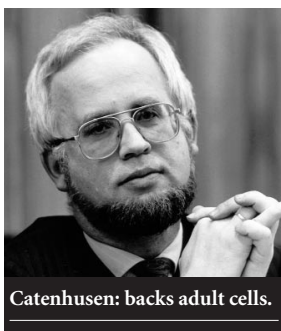
... but Germany remains unmoved

Quirin Schiermeier, Munich

Britain's support for research on cloned stem cells from human embryos (see above) presents Germany with a dilemma. Although neither politicians nor research organizations are seeking changes to Germany's restrictive embryo-protection law, both know the implications of such research taking place in a nearby country.

They are also relaxed at the prospect, however. For if British research shows that scientific techniques banned in Germany can lead to new medical treatments, Germany's attitude may no longer be easy to justify, they say.

But not everyone is convinced of the medical potential of embryonic stem cells. "Medical breakthroughs are not in sight," says Wolf-Michael Catenhusen,



Catenhusen: backs adult cells.

state secretary for research in the federal science ministry. "Our primary goal therefore remains to develop new therapies by using human adult stem cells."

The Deutsche Forschungsgemeinschaft (DFG), which funds most German university research, already supports work on adult stem cells under one of its priority

programmes. Like the government and the German Chambers of Physicians, the DFG sees no need to relax the embryo-protection law, aware of opposition from ethicists and theologians and widespread public hostility to the cloning of human embryonic stem cells.

"The mere expectation of new medical treatments is an insufficient justification," says Dietmar Mieth, head of the Centre for Ethics in Science at the University of Tübingen.

Others disagree. Given that both abortion and research on aborted fetuses are legal in Germany, "it is absurd that protection of embryos in culture is stronger than of those *in utero*," says Davor Solter, director of the Max Planck Institute for Immunobiology in Freiburg. ■

Green Bank dish may be last of the giants

Colin Macilwain, Washington

The largest steerable, single-dish radio-telescope ever built will be dedicated this week at Green Bank, West Virginia, in a ceremony to be attended by astronomers, politicians and research agency chiefs.

But the gargantuan Green Bank Telescope (GBT) — which is already six years late and tens of millions of dollars over budget — may mark the end of an era, as astronomy migrates from single-dish radiotelescopes to arrays of smaller instruments.

The 110-metre-diameter GBT has a unique reflective dish that sends radio waves from its field of observation to a receiving tower. This tower is offset from the dish itself, and so does not obstruct the telescope's field of view. The dish consists of an array of 2,000 panels, designed to move in real time to compensate for thermal or gravitational flexing of the structure and so keep the instrument sharply focused.

"It is an extraordinary instrument — even if we've been waiting for it a little bit longer than we had hoped," says Joseph Taylor, an astronomer at Princeton University in New Jersey, who won the 1993 Nobel prize for physics.

For Taylor and other pulsar astronomers, the GBT will offer an opportunity to conduct low-frequency observations of great sensitivity as soon as observation time on it becomes available, probably next spring.

But even as Senator Robert Byrd (Democrat, West Virginia) declares the telescope open, questions of finance remain unanswered. Comsat, the contractor responsible for the telescope's construction, has yet to hand it over to the National Radio Astronomy Observatory (NRAO), the arm of the National Science Foundation (NSF) that will operate the instrument.

Associated Universities Incorporated (AUI), the contractor that operates the NRAO for the NSF, is locked in dispute with Comsat over who should pay the extra costs that arose during the project's ten-year construction (see *Nature* **384**, 505; 1996).

Senator Byrd, who earmarked \$75 million of the NSF's budget to build the telescope in 1989, is thought to have favoured a dedication ceremony before November, when he stands for election to an eighth term in the US Senate.

The NRAO expects to take control of the telescope late next month. "I'm extremely hopeful that we'll have first light by the end of the year," says Phil Jewell, site director at Green Bank.

The GBT is of potential use to radio-astronomers interested in everything from star formation to astrochemistry. But although pulsar astronomers are likely to start their observations next spring, the GBT



Big spender: the Green Bank Telescope was completed six years late and millions over budget.

will only come into its own for many other astronomers if it can perform at higher radio frequencies of around 100 gigahertz, corresponding to wavelengths of 3 millimetres. At these frequencies the telescope will, for example, be able to detect a number of molecular tracers and so provide new insight into the early behaviour of young stars.

But high-frequency work will require the successful commissioning of the instrument's laser-based metrology system. This is supposed to control the position of its panels to within 25 micrometres, using a feedback loop of 12 ground-based lasers and the 2,209 motorized actuators on which the panels are mounted.

"This has never been tried before," admits Jewell. "But all of the hardware has been built now, and the results of our ground-based tests are very encouraging. I am very confident it will work."

The commissioning process will be complicated by fiscal constraints at the NRAO, whose budget next year will be frozen at \$32 million — even assuming that the US Congress allows the NSF the budget it has requested for 2001.

The 2002 House Appropriations Bill for the NSF contains language encouraging the agency to provide more funds for astronomy facilities, but does not actually provide any extra money. Robert Dickman, head of radioastronomy facilities at the NSF, says the agency's budget allows NRAO to compete for additional funds in areas such as information technology. "All the stakeholders are disappointed at the amount of time it has taken" to complete the telescope, he adds.

The budget crunch at the NRAO might

get worse when the AUI/Comsat arbitration process is completed. The cost overruns that are currently under dispute amount to \$28 million, according to Paul Vanden Bout, director of the NRAO, although some sources put the number even higher. Arbitration will end in October and a ruling is expected some time in 2001 — possibly allocating some of the costs to the observatory.

Engineers at the Green Bank site are still wrestling with several challenges, such as devising and implementing a suitable maintenance programme for the telescope's surfaces, welds and actuators. Vanden Bout is optimistic that this can be done. "We know what we have to do," he says.

But some astronomers worry that the NRAO is more fully committed to other projects, such as the planned Atacama Large Millimeter Array (ALMA) in Chile and the expansion of the Very Large Array (VLA) in New Mexico.

Asked if the NRAO was committed to the GBT as central to its future, Riccardo Giacconi, president of Associated Universities Incorporated, said: "We've got three of four major facilities, including the VLA and ALMA. GBT is something that we are committed to turn from a technological success to a scientific success." He concedes, however, that this will be "a tough job".

It's also a job that may be the swansong for efforts to scale up the fully steerable, single-dish radiotelescope, according to Taylor. "The trade-offs between a single large telescope and an array are such that it is not very likely that we'll want to build something that is even larger and fully movable," he says. ■

♦ <http://www.gb.nrao.edu/>

MIKE BAILEY/NRAO/AUI

US regulation threat over business links...

Paul Smaglik, Washington

US clinical researchers and their institutions received a stern warning last week: improve voluntary management of financial conflicts of interest, or face strict federal regulation.

That warning, from Greg Koski, currently professor of anaesthesiology at Massachusetts General Hospital in Boston, provides a preview of how he will run the Office for Human Research Protections (OHRP) when he takes up his new post on 5 September.

Ever since Donna Shalala, secretary of the Department of Health and Human Services, announced last summer that she would place the office, then under the jurisdiction of the National Institutes of Health (NIH), directly under her authority, researchers have been speculating over whether the new supervision would be looser or tighter.

Koski provided some hints at an NIH conference on financial conflicts of interest. Before his address, at the end of the two-day meeting in Bethesda, Maryland, many researchers had spoken in favour of greater 'guidance' — a euphemism for non-binding policy suggestions.

But Koski hinted that stricter measures might be on the way if this failed to work. "If guidance is not effective, then rules, regulations and laws will follow," he said.

Such measures may already be on the cards. Koski said that conflicts of interest in clinical research have become "pervasive". Physicians routinely get bonuses for encouraging their patients to become research subjects; academic researchers involved with clinical trials are often rewarded with equity in biotech companies; and paid consultancies by pharmaceutical companies to biomedical faculty members are on the rise.

Koski said he would avoid offering specific policy proposals to address those trends until he starts the job. But he sketched out a few ideas. For example, he suggested that universities might need new committees to adjudicate on conflicts of interest, as current institutional review boards (IRBs) could not be the "sole protector" against them.

And while several researchers spoke of the need for clinical scientists, universities and members of IRBs to make their financial interests public, Koski said that such measures are only a starting point. "Disclosure is not enough," he declared.

Several institutions detailed their approaches after being praised at the meeting by Ruth Kirschstein, NIH acting director, for representing 'best practices' in managing conflict of interest. Washington University in St Louis, for example, has strict rules: if a researcher has stock in a company, that company cannot fund the researcher's work.

Johns Hopkins University Medical School in Baltimore, Maryland, has, in addition

to an IRB, a conflict-of-interest panel that rules whether a researcher should put stock in a company that has interests in his or her research into escrow, or sell it off.

Both schools also scrutinize their own stock, and examine how their intellectual property could affect the perception of clinical trials. Any intellectual property owned by the university — such as a research tool used to perform an experiment — has to be described in the informed-consent forms that a patient must sign before entering a clinical trial.

Speakers pointed out that the appearance of conflict of interest is as important as evidence of wrongdoing. A perceived conflict of interest dogged James Wilson, director of the University of Pennsylvania Institute of Human Gene Therapy in Philadelphia, after the death of Jesse Gelsinger in a gene-therapy clinical trial last autumn.

Wilson had a large interest in a biotech company that produced viral vectors used for therapeutic gene delivery. Although an outside investigation found no wrongdoing, it concluded that Wilson's holdings caused an appearance of a financial conflict of interest (see *Nature* 405, 497; 2000). The university ended clinical trials conducted through the institute after the investigation.

Koski says the Pennsylvania incident galvanized interest in research conflicts of interest. Demands are already being heard in Congress, for example, for investigators to be



required to give more detailed financial information on informed-consent forms.

Opinions were mixed about the need for more federal regulations governing conflicts of interest. Michele Russell-Einhorn, director of regulatory affairs at OHRP, speaking for a group of scientists who met during the conference to discuss possible ways to manage them, said that "there is a consensus that there should not be regulations".

But Abbey Meyers, president of the National Organization for Rare Disorders, said that patient groups such as her own would like to see more protection against conflicts of interest. Regulation, rather than voluntary compliance, is the best way to ensure this, she said. ■

...as government seeks more ethics training

A proposal by the US federal government that all research staff — from managers to research assistants — should be required to take courses in the responsible conduct of research has met with a mixed reaction from organizations representing academic researchers.

Mary J. C. Hendrix, president of the Federation of American Societies for Experimental Biology (FASEB), says that the proposal for mandatory courses on topics such as biomedical ethics and conflict of interest is "overly broad". Last year a report commissioned by the NIH noted that the researchers it funded faced too many regulations (see *Nature* 398, 180; 1999).

But the Association of American Universities supports broad training in research conduct, Mark Brenner, vice



Koski: wants broader impact.

chancellor for research and graduate education at Indiana University told an NIH conference on conflict of interest last week (see above).

The requirements have been proposed by the Department of Health and Human Services' Office of Research Integrity, as a revision

of Public Health Service (PHS) guidelines for education in the "responsible conduct of research". FASEB has been pushing to extend the 21 August deadline for responses to the guidelines.

Hendrix agrees that students and trainees should take courses in research conduct, but says requiring everyone operating under a PHS grant to do so would be expensive and time-consuming.

Greg Koski, incoming director of the Office for Human Research Protections, supports broader training in the ethics of research. "The education needs to go well beyond the investigator," he said during the conference. **P. S.**

► <http://ori.dhhs.gov/TheRCRPolicy.htm>
 ► <http://www.faseb.org/opar/ltr/oriltr.html>

US and Vietnam plan joint dioxin research

Rex Dalton, Monterey, California

In what some see as a unique research opportunity and others as an ironic footnote to history, the first US–Vietnamese research programme on dioxin pollution entered the planning stage last week.

A multidisciplinary panel of researchers assembled by the National Institute of Environmental Health Sciences (NIEHS) met in an attempt to identify research gaps that could be filled by bilateral studies.

The panel argues that Vietnam offers an unprecedented chance to conduct important research on human exposure to dioxin.

“This is it,” testified Michael DeVito, a toxicologist who studies dioxin for the Environmental Protection Agency and has visited Vietnam. “We may never again have a population exposed to this level.”

It is more than 30 years since the United States first sprayed the defoliant Agent Orange over the forests of Vietnam. The herbicide was contaminated with one of the most toxic dioxins — TCDD, often called simply dioxin — exposing large numbers of Vietnamese to the chemical.

Suggested topics for the joint programme, funding for which has been approved by Congress on condition that Vietnam also provides support, include a study on the carcinogenic effects of dioxin, an examination of developmental diseases (neurological and growth) in children exposed to the chemical, and new methods for assaying and analysing residual contamination.

NIEHS officials will use testimony from last week’s meeting to draw up a research plan for the delicate discussions with their Vietnamese counterparts. No funding or starting point have yet been set for the research. But US scientists on the panel said it should be



Apocalypse continuing: Vietnam is still suffering from the Agent Orange dropped on it in the 1960s.

done soon to capture the human and environmental impact of dioxin while worthwhile research results can still be obtained.

In return for its contribution to US studies of the human and environmental impact of the wartime spraying, Vietnam, which has limited research funds and lacks sophisticated equipment, is expected to seek US assistance with its pollution problems.

Christopher Portier, acting director of environmental toxicology for the NIEHS, hopes a US–Vietnamese meeting will take place within a few months in a neutral nation. “We are trying to get the governments to meet to set up a collaborative research effort,” he says. “Then we can go to academic scientists in both countries for proposals, which will be considered on a competitive basis.”

There were no representatives of the Vietnamese government at last week’s meeting,

which was held in Monterey following the 20th International Symposium on Halogenated Environmental Organic Pollutants and POPS (persistent organic pollutants).

But one Vietnamese researcher, physician Le Cao Dai, executive director of the Agent Orange Victims Fund set up in Hanoi by the Vietnam Red Cross, attended both meetings to present research and offer suggestions for scientific projects. Dai said that Vietnamese scientists are keen to carry out joint research.

The proposed project will have to overcome a bitter history of war, devastating environmental pollution resulting from the spraying, and decades of diplomatic acrimony. US military forces used Agent Orange in Vietnam between 1962 and 1970. The herbicide was used to kill vegetation around military installations and along transportation routes used by Vietnamese forces.

The land remains unforested, and many people still face exposure from residual dioxin in soil, water and animal fats, for example in fish and ducks. Panel member Linda Schwartz, a Yale University nursing researcher and Vietnam veteran, said that parts of Vietnam resemble “a huge, environmental Superfund [toxic dump] site”.

Five years ago, an attempt at bilateral dioxin studies foundered when US scientists and Vietnamese officials fell out over what pollutants to study and how specimens should be handled. Although new diplomatic links have helped overcome these disagreements, some new problems have emerged.

For instance, the ownership of any intellectual property from the project has been a sticking point in negotiations, says physician Michael Linnan, US health attaché to Vietnam. Many Vietnamese are concerned about US scientists capitalizing on research into the wartime spraying of their country.

♦ <http://www.niehs.nih.gov/oc/factsheets/dioxin.htm>

Biodiversity cash aimed at hotspots

A \$75 million fund to preserve the world’s most endangered biodiversity locations was announced this week, with plans to double the amount soon.

The World Bank, Conservation International and the Global Environment Facility (part of the United Nations Development Programme) have each pledged \$5 million a year for five years for rapid intervention where biodiversity is threatened. The agencies hope other groups will donate at least another \$75 million.

During the first year of operation, the Critical Ecosystem Partnership Fund plans to focus on maintaining biodiversity in Madagascar, West Africa and the Tropical Andes in South America.

The fund will provide grants to manage areas of high biodiversity (known as hotspots), resolve conflicts with industries that destroy natural regions, and create partnerships between private firms and preservation agencies to address biodiversity issues. Planning, training and support services will also be funded.

“This is a new source of money exclusively for local groups whose work is central to protecting the biodiversity hotspots,” says Peter A. Seligmann, chief executive of Conservation International, a Washington-based agency formed to conserve biodiversity (see *Nature* 403, 853–858; 2000).

♦ <http://www.cepf.net>

R.D.

Iceland's doctors rebuffed in health data row

Alison Abbott

The Icelandic Medical Association (IMA) has failed to reach agreement with the genomics company deCODE Genetics over Iceland's new national health sector database, which is to be run by the company. The dispute centres over the plan for automatic inclusion of patients' health records in the anonymized database.

Opponents are concerned that deCODE, which launched itself on the Nasdaq stock exchange last month, is painting an excessively optimistic picture of the conflict to shareholders.

The new system will collect and store information from the Icelandic health service. Under the present rules, patients would be presumed to have given their consent to inclusion even if they had not been consulted.

The Icelandic government approved plans for the database in 1998. In January this year it granted deCODE a 12-year licence to operate the database and market — in encrypted form — the information it contains together with the company's own genomics data of the Icelandic population.

A law passed in 1998 requires health records to go into the database unless an individual actively opts out. But the IMA insists that transfer should be based on informed,

rather than presumed, consent — the issue should be discussed with each patient.

Six months ago, the two sides embarked on discussions to resolve the issue. But compromise has proved impossible, and a joint press release was issued last week reporting that talks "had been concluded".

The IMA is upset that deCODE has issued an English translation of the press release — now posted on the Nasdaq website — giving a more positive slant to the discussions than the Icelandic original had. The Icelandic release, for example, says that the talks "covered the content and have been to the point", whereas deCODE's English version says the discussions "have been both positive and productive".

Sigurbjörn Sveinsson, chairman of the IMA, says that deCODE "did not seek our approval of their English version, and there is obvious inconsistency in the two versions".

But Kári Stefánsson, deCODE's chief executive officer, says the company does not need the IMA's approval to proceed with the database on the basis of presumed consent. "The law is quite clear," he says, adding that he is still keen to make peace with the IMA.

Most Icelandic doctors oppose presumed consent, and many say they will not hand over records that are not already in hospital



It's the law: Iceland's parliament (above) has approved 'presumed consent'.

databases. "I don't feel comfortable with handing over data when we don't know how it will be used in the future," says Sigurdur Björnsson, an oncologist on the IMA board.

Stefánsson says that only a few records remain outside public institutions, and the project does not require support from all the doctors, as the database does not need to represent 100% of the population to be useful. ■

♦ <http://www.decode.com>

♦ <http://www.mannvernd.is/english/index.html>

Astronomers reveal the secrets of the name game

Steve Nadis

Perhaps it was inevitable. Features on the surface of Eros, the near-Earth asteroid now being observed by the NEAR Shoemaker spacecraft, have officially been given the names of famous lovers in myth and literature, such as Orpheus and Eurydice, Don Juan, Cupid and Lolita.

Assigning such names is the job of the working group for planetary-system nomenclature of the International Astronomical Union (IAU), which convened this month in Manchester, England, at a meeting of the IAU general assembly. The names Bill and Monica are said to have been summarily rejected.

This year the working group has selected more than 200 names for the satellites of Uranus and for craters and other features seen on the Moon, planets and asteroids.

The system is intended to make life easier for researchers, rather than commemorate people, explains Brian Marsden, a member of the working group who is based at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts.

Yet commemoration plays an important role. A crater on the moon was named after



Commemorated: Shoemaker (left) and Sagan.

the planetary scientist Eugene Shoemaker, for example, and one on Mars was named after the astronomer Carl Sagan — both of whom became eligible, according to IAU rules, three years after their deaths.

But the desire to honour Shoemaker and Sagan could create problems, says Marsden. "There is a tradition that you don't use the same name for different astronomical bodies, but these names have been mentioned for other objects as well."

Another point of contention was raised by some of the younger astronomers, who complained that most members of the working group have held their posts since

the 1970s. "There's a feeling that people should not do more than three [three-year] terms, so that others won't feel left out," says Marsden, who is starting his third term and is prepared to retire after the next IAU general assembly meeting in 2003.

One change introduced this year establishes an appeal process, allowing any IAU member to object to a new name within three years of its announcement. No such challenges have yet been raised, but a Japanese amateur astronomer did ask the working group to reconsider the name for an asteroid he discovered. His suggestion had been rejected because it was the name of the company he worked for, but the committee reversed its decision, and the name, Denso, was approved.

Most names are drawn from history and literature, rather than the corporate world. Names for the moons of Uranus — Caliban, Sycorax, Setebos, Stephano and Prospero — were taken from Shakespeare's play *The Tempest*. The last three names replace the temporary designations S/1999U1, U2 and U3. "I'm perfectly happy to use numbers, which I find easier to remember," Marsden says. "But the public likes to have names." ■

Human Genome Project makes its mark in Germany

Munich News of the draft human genome sequence has not been slow to reach the German public. According to a survey carried out last month, 84% of Germans say that they are aware of the achievement, a surprisingly high proportion for a country in which the circulation of news about research tends to be slow, and public understanding of science is often said to be low.

The imagination of many Germans seems to have been fired by a development that many believe will be epoch-making. More than half of those who know about the completion of the project said that it would be "more significant than the first landing on the Moon".

Many Germans remain sceptical. Only a third were optimistic about the project's potential benefits, whereas 44% were afraid of the consequences. But in eastern Germany, more expressed hope (39%) than fear (37%).

Global warming concerns over melting North Pole

London The icecap at the North Pole is melting, says a scientist who found a mile-wide stretch of open ocean on a recent trip there. James McCarthy, director of the Museum of Comparative Zoology at Harvard University and co-chair of a working group of the Inter-governmental Panel on Climate Change, told *The New York Times* he had expected to find an icecap of between six and nine feet thick. But on a trip to the pole on a Russian ice-breaker, the *Yamal*, earlier in August, he found only a thin crust of ice.

Although the discovery is being seen as further evidence of the existence of global warming, Peter Wadhams, reader in polar studies at Scott Polar Research Institute at the University of Cambridge, says it is common to find open water at the North Pole in summer. But he also says there was an unusually large amount of open water this year — part of a recorded thinning and retreating of sea ice. "Mean ice thickness in late summer is now 40% less than it was 20–30 years ago," he says.

Siena banks on local science research

Siena, Italy The world's oldest banking institution — Monte dei Paschi di Siena, founded in 1472 — is setting up a fund to support local research. The fund, due to be launched next year with an annual budget of 10 billion lire (\$10.4 million), will be administered by the bank's charitable foundation.

Rino Rappuoli, director of IRIS, a



Hot tip: forget the races, Siena is backing science.

vaccine-development company owned by California-based biotechnology company Chiron, and a member of the foundation's scientific advisory committee, says the money will probably be used to extend Siena's strong base in medical research. His committee has also suggested that it should fund the development of genetic tests on new vine stock, an important part of the local economy.

Genetic animal experiments on the rise in Britain

London The number of animals used for genetic experiments in the United Kingdom rose by 14% to 512,000 last year — fuelling speculation that a 25-year decline in the use of animals in research is ending. The figures, released by the Home Office last week, drew a warning from the Royal Society for the Prevention of Cruelty to Animals (RSPCA) that the use of genetically engineered animals might be getting out of control.

The RSPCA wants more scrutiny of these experiments. But researchers are already complaining that the administrative burden is too great (*Nature* 404, 529–530 & 405, 725; 2000). The Home Office is reviewing the system of processing applications for research projects.

Putin to head Russian science policy council

Moscow Russian president Vladimir Putin is to head a new council dealing with science policy. His announcement came at the end of a meeting last week with members of the Russian Academy of Sciences, which addressed issues ranging from staffing problems to the Internet's impact on Russian science.

Arguing for a more positive attitude to exploiting scientific discoveries, Putin said that "little has changed since the times when inventions were made in Russia but applications for them were found elsewhere". He also admitted that more government support for basic research was needed,

adding that the improvement over the past 18 months was not enough.

MediaLabEurope ready to 'reinvent the future'

Dublin MediaLabEurope — a research and education centre established in Dublin, Ireland, by the Massachusetts Institute of Technology Media Lab — was officially opened at a ceremony last month (see *Nature* 401, 836; 1999). Speaking at the two-day event, Bertie Ahern, the Irish prime minister, said that the laboratory's mission is to "reinvent the future".

The lab aims to train researchers, inventors and artists, in Ireland and throughout Europe, in new communications and multimedia technology. The Irish government has provided IR£28 million (US\$32.3 million) in initial funding to support the venture, as well as a laboratory building in central Dublin at the site of a former brewery.

IBM turns its attention to life sciences

Washington The computer giant IBM is investing \$100 million to develop computational biology tools, and is expanding its efforts in biology by forming a life sciences division. Both moves are based on the company's estimate that the market for information technology in the life sciences will rise from \$3.5 billion today to more than \$9 billion by 2003.

The company predicts that much of the need for computing in biology will be driven by the Human Genome Project, proteomics projects and the pharmaceutical research that are likely to spin off from both. Last December, IBM committed \$100 million to building a supercomputer designed to help scientists understand the mechanisms behind protein folding (see *Nature* 402, 705; 1999).

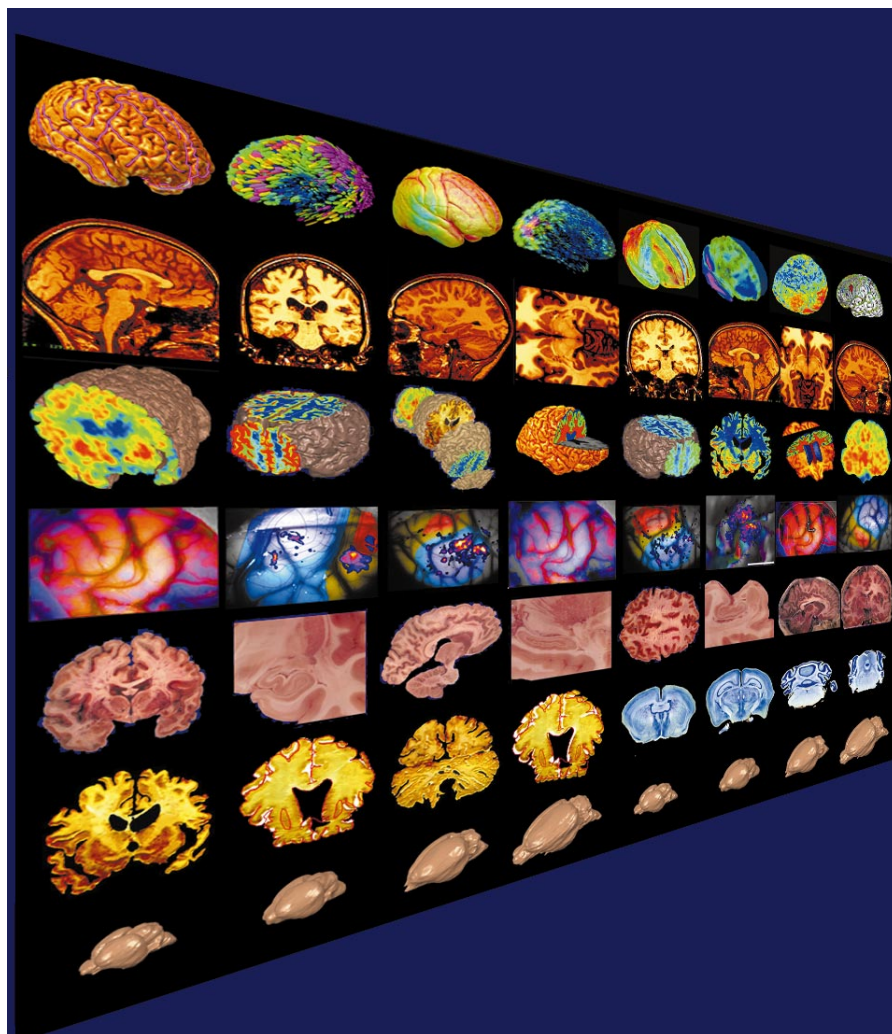
Brain-dead organ donors 'should be anaesthetized'

London Anaesthetic should be routinely given to brain-dead organ donors, according to an editorial in *Anaesthesia*, the journal of Britain's Royal College of Anaesthetists. Whether or not brainstem-dead patients experience pain when organs are removed remains a controversial issue among anaesthetists.

Patients are already given muscle-paralysing drugs to prevent involuntary responses to surgery, but anaesthesia is not required, and guidelines issued last year by the Intensive Care Society stated that this was not necessary. The founder of the British Organ Donor Society has criticized the journal's comments as unsettling for relatives of the patients involved.

Databasing the brain

Progress in neuroscience might be faster if researchers shared their results in a network of databases. But the technical challenges are huge, and reaching a consensus on what to archive won't be easy, says Marina Chicurel.



ANDREW LEE, PAUL THOMPSON & ARTHUR TOGA

Stephen Kuffler probably did not realize what he was starting when he established the Department of Neurobiology at Harvard Medical School in 1966. Kuffler helped found the modern discipline of neuroscience by bringing together physiologists, biochemists and anatomists to focus their efforts on the nervous system.

Today, more than 50,000 neuroscientists worldwide study everything from individual molecules to complex behaviours in species from nematode worms to humans. Generating enough data to fill more than 300 journals, they have created one of the largest, most unwieldy datasets in science. Several prominent neuroscientists are now arguing that the time has come to tame this monster. They believe that progress could be boosted by creating interoperable databases, allowing researchers to share their results and make links between data from labs around the world. "It's a topic that needs to be spotlighted," says Dennis Choi of the Washington University School of Medicine in St Louis, who is president of the Society for Neuroscience.

But there are three big obstacles. First, reaching a consensus on what is worth including in databases. Second, the techni-

cal difficulty of collating and relating such disparate types of information. And third, the reluctance of researchers who have traditionally guarded their results jealously to embrace data sharing. "It's a data-hugging community," observes Michael Arbib, director of the Brain Project at the University of Southern California in Los Angeles, which is developing a variety of neuroscience databases.

Where to begin?

Bioinformatics is most advanced in genomics and proteomics, where DNA and protein sequences are routinely archived in public databases. Indeed, many journals will only consider papers if authors make their sequences public. But linear sequence data are easy to store in this way.

"Technically, the genome and protein databases are relatively trivial," says Stephen Koslow, director of the Office of Neuroinformatics at the National Institute of Mental Health in Bethesda, Maryland. "Neuroscience data are much more complex." But in the latest issue of *Nature Neuroscience*, Koslow argues that the technical obstacles can be overcome, and makes a plea for more sharing of data¹.

Koslow and others point to projects that are showing the way forward. Many of these are supported by the Human Brain Project, a multi-agency US government initiative established in 1993 to support research into databases and related tools for neuroscientists. Bioinformatics enthusiasts note the scientific advances made possible by such databases (see 'Making the connections', opposite). And they are encouraged by the Organisation for Economic Cooperation and Development's (OECD's) recent acceptance of a proposal to set up a working group on neuroinformatics, chaired by Koslow, which aims to establish guidelines for neuroscience databases and software, and provide an Internet portal for their dissemination².

But shovelling data into a database is only useful if they can be organized in meaningful ways. And in some areas, it is still unclear whether neuroscientists know enough to sort the wheat from the chaff. Geneticists agreed long ago on the value of storing reproducibly generated DNA sequences, but not images of their sequencing gels. Many of neuroscience's subdisciplines have yet to reach a consensus on what is worth storing. In addition, techniques that neuroscientists use to generate and analyse data are often not



Stephen Koslow.

Technically, the genome and protein databases are relatively trivial. Neuroscience data are much more complex.

standardized. "Part of this process must be to confront the devils early in the game," says Choi. "The devils include the possibility of prejudice overly influencing future thought, essentially by controlling the shape of the database or the priority given to certain data over others."

Most database developers agree that flexibility is the key. Daniel Gardner of the Weill Medical College of Cornell University in New York is developing the Cortical Neuron Net Database, a repository for neurophysiological recordings from the brain's cortex. Trying to ensure that the database keeps pace with the field, Gardner's team is developing a hierarchical classification scheme that allows for new categories to be introduced without making earlier entries obsolete or restricting searches. "It's never economically possible to go back and re-classify old databases," says Bruce Schatz, an information scientist at the University of Illinois at Urbana-Champaign. "It just costs too much."

Lack of standardization is a persistent problem, and the confused nomenclature of neuroanatomy is a particular bugbear. Given that the same brain region can go by the name of the caudate nucleus or nucleus caudatus, or may be referred to as part of a larger structure known as the basal ganglia, only experts can reliably sort the data. Worse still, different neuroanatomists sometimes use the same names to refer to different structures. "The actual business of data collation is a hugely, hugely onerous thing," says Malcolm Young, a systems neuroscientist at the University of Newcastle-upon-Tyne in England. "It's more onerous in my experience than doing experimental neuroscience."

Some scientists are turning to computers for help. Researchers led by Rolf Kötter at the Heinrich Heine University in Düsseldorf, Germany, have developed an algorithm called Objective Relational Transformation (ORT)³ for their database of cortical connectivity in the macaque brain, known as CoCoMac. This algorithm transforms neuroanatomical data from one representational format, or mapping scheme, into another. By applying ORT to data in their original formats, individual database users can convert the data into any format they wish to work with. "Besides reducing costs, the approach is more objective, reproducible, transparent and correctable than human labour," says Kötter.

There is also more disagreement over the

interpretation of neuroscience data than there is for DNA sequences. Gordon Shepherd and his colleagues at Yale University in New Haven, Connecticut, are working on this problem. Their SenseLab database integrates information on the olfactory system, including the sequences of olfactory receptor proteins and the electrical properties of the neurons involved in processing smells.

SenseLab flags controversial data through annotations that describe conflicting studies and point users to the primary literature.

Other database architects are designing filters to let users decide how to weight the data. At the University of Southern California, Gully Burns is constructing a database on neural connectivity in the rat brain called NeuroScholar. Future users will be able to score studies based on attributes such as the journal they appeared in or the techniques the investigators used. They will then be able to combine the weighted data from several studies to assess the strengths of hypotheses they might wish to test.

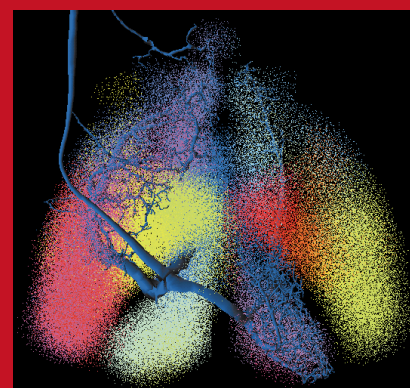
Other researchers agree that approaches such as this, which allow users to impose their own organizational schemes on the data are, in theory, the best solution. James

Making the connections

Why bother to build neuroscience databases? Ask a bioinformatics enthusiast, and he or she will fire back a handful of success stories that provide a glimpse of the future.

The International Consortium for Brain Mapping, which has built a database of brain images from 7,000 people⁸, is one example. As human brains vary with age and between individuals, it is impossible to create an absolute atlas of structure and function. But by using sophisticated algorithms to align a large numbers of images, the consortium has generated atlases that show the probability that any given region of the brain is involved in the performance of a certain task, or is disrupted by a particular disease. Atlases that highlight regions commonly affected by Alzheimer's disease and stroke, for example, are becoming powerful tools for tracking the effects of candidate therapies in clinical trials. John Mazziotta of the University of California at Los Angeles, who heads the consortium, argues that the database promises to become even more powerful as researchers begin to match up the brain images with additional information such as genomic, neurochemical and behavioural data.

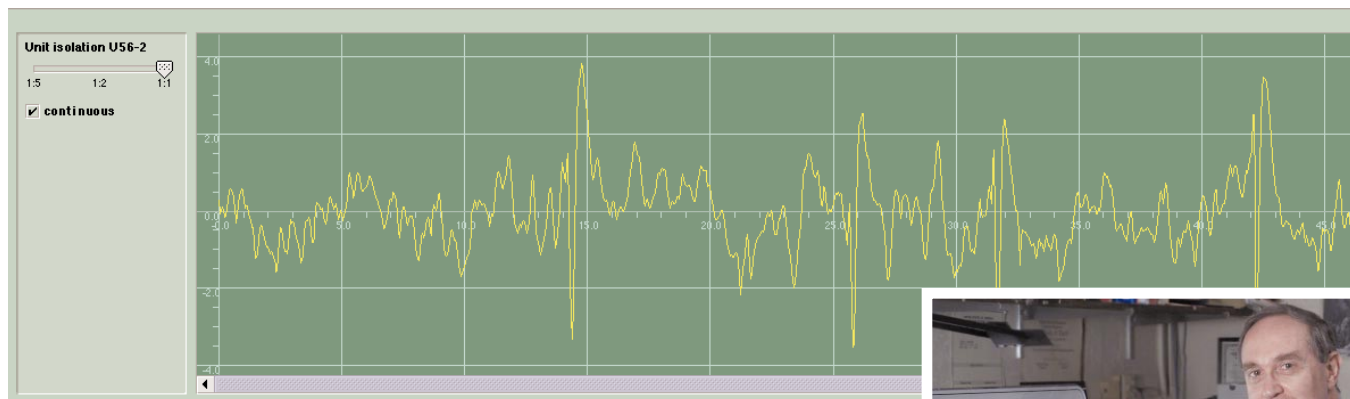
Meanwhile, Malcolm Young at the University of Newcastle-upon-Tyne is gaining new insight into how the brain is hooked up by mathematically sorting through thousands of anatomical studies. It started in 1990, when Young set eyes on a global diagram of the hundreds of pathways that interconnect the visual cortex, drawn up by David Van Essen of the Washington University School of Medicine in St Louis⁹. He soon realized that by applying mathematical analyses to the collated data, he could mine it for valuable nuggets of information¹⁰. Since then, Young has developed additional methods, successfully predicted new relationships between structure and function in the brain, and is helping to spawn a new field of research in which the analysis of anatomical



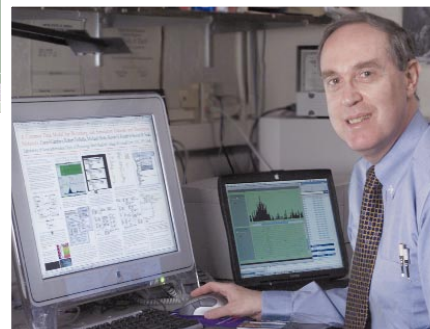
Plugged in: a NeuroSys image of a neuron's inputs. Each colour represents the sensing of air motion from a particular direction.

data is helping researchers to localize neurons with specific electrophysiological properties¹¹.

Bringing together electrophysiological and anatomical data on more than 200 neurons, Gwen Jacobs at Montana State University in Bozeman is discovering how networks of neurons encode sensory information. Neuroscientists are often limited to studying small numbers of cells at a time, rarely seeing the big picture of how larger neural networks transmit complex messages. But by focusing on a simple system and setting up NeuroSys, a network of databases that allows her to integrate data from multiple neurons, Jacobs is overcoming this limitation. "It came out of the frustration of trying to hold ten different variables in your head at the same time," she says. Jacobs has collected data on the behaviour and shape of neurons that are involved in detecting air motion in crickets. Plugging her findings into NeuroSys, she has discovered that the coding system used to specify the direction that a puff of air has come from seems to depend on the branching patterns of a small set of key neurons¹².



Virtual oscilloscope: Gardner's invention lets researchers view neurophysiological recordings.



experts such as Schatz say that computer power and Internet bandwidth should soon cease to be limiting factors.

That should assist neuroscientists in tackling perhaps the most difficult challenge of all: linking their various databases into a seamless federation. Ideally, says Schatz, users should be able to navigate between databanks without noticing borders, as if interacting with a single, cohesive database. But this has proved difficult, even in fields with much simpler datasets. It took years to connect the European Molecular Biology Laboratory's DNA sequence database with the SWISSPROT protein sequence database, because researchers could not agree on nomenclature.

Learning to share

Many database developers believe the best approach is to standardize the communication between databases, rather than standardizing the data they hold. One helpful software trick is to 'wrap' data with universal labels that describe their composition and format so they can be understood by different types of software, running on different computers — the HTML code that underlies web pages is one such wrapper. Gardner is developing an interface to link different databases using the Biophysical Description Markup Language, a wrapper that he designed specifically to label neuroscience data, based on a code called XML. Schatz believes another key tool will be concept-switching, a search strategy that relies on analysing the contextual relationships between phrases to identify underlying concepts^{5,6}. Concept-switching algorithms, for instance, would identify the term 'nucleus caudatus' used in one database as occurring in similar contexts as 'basal ganglia' in another, and link them together.

But the technical issues are not the only

Bower at the California Institute of Technology in Pasadena, for example, is developing model-based database systems whose organization relies almost exclusively on the user. Bower's GENESIS Database holds data that range from the characteristics of ion channels in cell membranes to the details of the connections between neurons. But these data only become usefully organized when a user creates a model of a neuron or network of neurons.

Super models

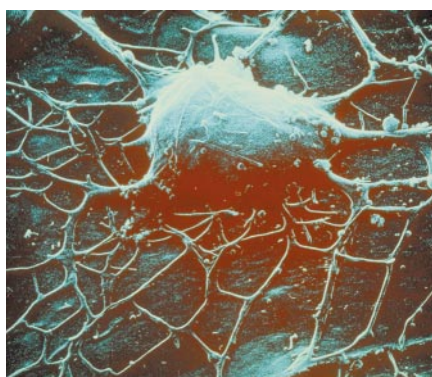
For example, a user may devise a model of a Purkinje cell — a type of neuron found in the cerebellum — to predict how electrical signals from its projections, or dendrites, will be transmitted to the cell body. The model will then sort relevant data from the database, and help the researcher spot inconsistencies. If the model is not compatible with known information, or if data are contradictory, the model will alert the researcher to the problem. And because the models are formal representations of hypotheses, they provide a way for neuroscientists to communicate and compare their ideas quantitatively.

More problems are posed by data that vary in three dimensions. Like a crumpled beach ball, the human brain's surface is covered with folds, serving as natural landmarks for neuroanatomists comparing results. But

the distribution of folds varies between individuals, as does their location relative to functional brain areas. Some researchers have resorted to shrinking or expanding their images to align them with different-sized brains. But brains often vary in more complicated ways than size. So several researchers have developed more sophisticated warping algorithms. David Van Essen of the Washington University School of Medicine, for example, has created algorithms that flatten the surface of the brain. These reduce the alignment problem to two dimensions, and have been incorporated into Van Essen's software toolkit and database of cortical structure and function, the Surface Management System.

Neurophysiologists, meanwhile, generate large amounts of time-series data as they record the electrical behaviour of neurons. Michael Gabriel and his colleagues at the University of Illinois have developed a time series data protocol which has been incorporated into the Neuronal Time Series Analysis Workbench, a set of informatics tools designed to help researchers who work with such data. Like Kötter's ORT, the protocol is associated with translational filters that allow users to work with data in a variety of formats. Another tool for neurophysiologists is a virtual oscilloscope, devised by Gardner at Cornell⁴. This allows electrophysiologists to see extended datasets — as opposed to the small, static excerpts that appear in published articles — as well as artificial traces predicted from simulations.

Neuroscientists often have to wrestle with huge computer files. An image of a single section of monkey brain can take up 80 megabytes, says Edward Jones of the University of California at Davis, who is building brain atlases of such high resolution that they include individual neurons. A typical atlas of a monkey brain includes 1,500 sections, so his files often fill several gigabytes. Functional magnetic resonance imaging (fMRI) datasets pose similar problems. Having access to servers that can handle such monstrous files and ship them across the Internet is problematic. But computer



Networking: databases allow users to model the ways that neurons link together.

ones that need to be ironed out. In a highly competitive field, in which data are often hard-won, many researchers are reluctant to release their data to be exploited by others. Some neuroscientists, such as Peter Fox at the University of Texas Health Science Center in San Antonio, also argue that the complexity of the data poses problems. When Fox sent primary brain-imaging data to a collaborating lab, he says he had to be in almost continuous communication to explain how he had calibrated his machines, subtracted noise and displayed the data. If primary data were freely available, Fox fears that some researchers might be led to the wrong conclusions. "The submission of truly raw data would lead to a nightmare of misinterpretation," he says.

The sensitivity of the data-sharing issue was underlined last month, with the opening of a new repository for fMRI data at Dartmouth College in Hanover, New Hampshire. A letter sent to leading researchers in the field announcing the National fMRI Data Center by its director Michael Gazzaniga raised fears that, in the future, journals might require primary data relating to papers on fMRI to be deposited in the database⁷. Gazzaniga sent the letter in his capacity as editor of the *Journal of Cognitive Neuroscience*, which had decided to adopt such a policy. That prompted several dozen fMRI specialists to sign a letter arguing against mandatory data submission, which was sent to the data centre's financial backers and to the editors of 14 leading journals.

Although it may not be universally applicable, Yale's Shepherd has devised a way of encouraging neuroscientists to share



Flat pack: Van Essen irons out the brain's wrinkles to make mapping and comparisons easier.

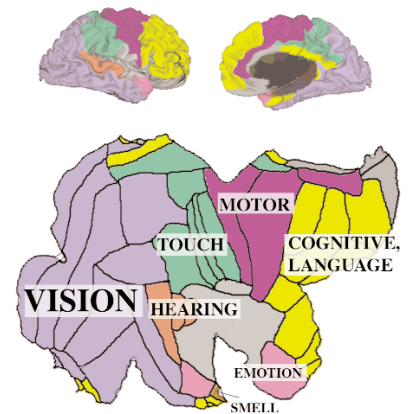
without losing control of their data. The olfactory receptor database within Shepherd's SenseLab allows users to deposit unpublished sequences, which are then kept hidden from other users' view. When a search for sequence homologues finds a match with one of these unpublished sequences, the database provides the searcher with the contact information for the researchers who submitted the sequences. In this way, researchers can avoid revealing their primary data, yet benefit from identifying potential collaborators.

The task ahead

Koslow accepts that it may take some time before neuroscientists adjust to a culture of data sharing. But he argues that the complexity of data do not pose insurmountable problems. For example, he says, details of the experimental conditions and variables needed to interpret primary data could be incorporated into databases. Koslow believes that, ultimately, neuroscientists must be allowed to come to a consensus on the desirability of data sharing, and that attempts by funding agencies or journals to force the pace could be counterproductive.

When it comes to building the basic infrastructure, however, Koslow is keen to push the pace. But experts disagree on how to move forward. Koslow wants to expand the existing approach of funding a wide variety of small projects, and then building links between those that prove most successful — and is pushing this model in the OECD working group. Schatz is in favour of more centralization. In his view, small projects based in individual labs, often relying on graduate students who are amateurs in informatics, are likely to fail. He thinks the big problems, such as how to form database federations, have to be tackled by experts working in dedicated informatics centres. Based on his experience in the early days of genomics, Schatz proposes getting professional informaticians to build a network of databases focusing on a particular organism and a set of well-studied brain regions. He

Functional Subdivisions of Human Cerebral Cortex



Van Essen & Drury, 1999

also believes the field needs definite targets against which to measure progress.

As the experts debate the way forward, some neuroscientists are sceptical about investing heavily in databases. Says Rodney Douglas, a computational neuroscientist at the Institute of Neuroinformatics* in Zurich: "Is it really a good thing for us to spend a lot of money on cataloguing everything in sight?" Douglas argues that the priority in neuroscience should be to understand how nervous systems deal with information. "The nervous system doesn't collect a lot of data, it collects relevant data, and I think we can learn from that," he says.

The best answer to the sceptics would be to demonstrate the practical advantages that databases can bring. But Koslow, who chairs the Human Brain Project's coordinating committee, concedes that the project's overall impact has so far been modest. The challenge for the bioinformaticians is to produce some tools that can make a real difference to working neuroscientists. "I became convinced early on of the need to build databases," says Van Essen. "But I must admit I didn't realize how challenging a task that would be."

Marina Chicurel is a freelance writer in Santa Cruz.

For an online debate on data sharing in neuroscience, see www.nature.com

Web links:

Human Brain Project

♦ <http://www.nimh.nih.gov/neuroinformatics/index.cfm>

University of Southern California Brain Project

♦ <http://www-hbp.usc.edu>

Cortical Neuron Net Database

♦ <http://cortex.med.cornell.edu>

CoCoMac

♦ <http://www.cocomac.org>

SenseLab

♦ <http://ycmi.med.yale.edu/senselab>

NeuroScholar

♦ http://www-hbp.usc.edu/Projects/neuroScholar_Connx.htm

GENESIS Database

♦ <http://www.bbb.caltech.edu/hbp>

Surface Management System

♦ <http://stp.wustl.edu/sums>

Neuronal Time Series Analysis Workbench

♦ <http://soma.npa.uiuc.edu/isnpa/isnpa.html>

International Brain Mapping Consortium

♦ <http://www.ionu.ucla.edu/ICBM/index.html>

NeuroSys

♦ <http://nervana.montana.edu/projects/neurosus>

1. Koslow, S. H. *Nature Neurosci.* **3**, 863–866 (2000).
2. Smaglik, P. *Nature* **405**, 603 (2000).
3. Stephan, K. E., Zilles, K. & Kötter, R. *Phil. Trans R. Soc. Lond. B* **355**, 37–54 (2000).
4. Gardner, D., Abato, M., Knuth, K. H., DeBellis, R. & Erde, S. M. *Phil. Trans R. Soc. Lond. B* (in the press).
5. Butler, D. *Nature* **405**, 112–115 (2000).
6. Schatz, B. R. *et al.* *IEEE Computer* **32**, 51–59 (1999).
7. Aldhous, P. *Nature* **406**, 445–446 (2000).
8. Mazziotta, J. C. *et al.* *Neuroimage* **2**, 89–101 (1995).
9. Felleman, D. J. & Van Essen, D. C. *Cereb. Cortex* **1**, 1–47 (1991).
10. Young, M. P. *Nature* **358**, 152–154 (1992).
11. Young, M. P. & Scannell, J. W. *Phil. Trans R. Soc. Lond. B* **355**, 3–6 (2000).
12. Jacobs, G. A. & Theunissen, F. E. *J. Neurosci.* **20**, 2934–2943 (2000).

*In the United States, the term neuroinformatics generally refers to the application of databases and related tools to neuroscience. In Europe, it is sometimes used to describe the discipline of computational neuroscience, which models the behaviour of neurons and neural networks.

Industry benefits from the public funding of intellectual curiosity

Sir — I was contacted recently by the *Los Angeles Times*, which had heard that the concept of expressed sequence tags (ESTs), essential to the human genome project, was developed in our laboratory. It is not generally known that we had described this development in 1983 in *Nature*¹, preceding by many years the adaptation of the technique to the genome project^{2,3}.

The genome project has important commercial applications. Still, I think it is essential for the community at large to appreciate that the development of the concept of ESTs and shotgun sequencing had nothing to do with commercial interests, but rather was motivated by our intellectual curiosity. The generosity and open-mindedness of the National Institutes of Health and Muscular Dystrophy Foundation enabled us to tackle the question with the spirit of 'let's just do it and see what happens'.

In 1982, my laboratory was in the Department of Biology at Massachusetts Institute of Technology. DNA sequencing was a relatively new method developed along two different and independent lines by Walter Gilbert at Harvard and Fred Sanger at the MRC Laboratory in Cambridge, UK. We had started with the Maxam–Gilbert chemical method in the late 1970s, but we gradually incorporated Sanger's powerful dideoxy approach in the 1980s — now used almost exclusively in automated instruments.

Most people then were interested in obtaining sequences of individual genes or of a complementary (c)DNA copy of a single messenger (m)RNA. This was viewed as an enormous undertaking, and typically involved determining a partial sequence of the desired gene product (a protein), then making a degenerate oligonucleotide for hybridization and primer extension or reverse transcription. Eventually, a gene would be cloned and sequenced. Often, however, the degenerate primers did not yield anything useful because they cross-hybridized to spurious sequences.

The question we asked was as follows: what if we simply did random, shotgun sequencing of all expressed genes associated with a specific tissue? If we then translated these sequences into all six reading frames, we should pick up fragmentary amino-acid sequences associated with the proteins of that specific tissue. These would include known and unknown proteins. We would have to use a single primer that bound to all expressed sequences — the mRNAs. Because eukaryote mRNAs have

'tails' of polyadenylic acid (poly A) at their ends, a single primer of oligo deoxy-thymidylate (oligo dT) could be used. With this primer, we could reverse-transcribe the mRNA pool and thereby pick up sequences near the ends of those mRNAs. These were the expressed sequence tags (although we did not call them that then).

At the time this idea seemed a bit far-fetched. The only way we could know whether we were on the right track was if among the sequences we obtained were some that clearly coded for the proteins specific to the tissue we had used for the preparation of the cDNA. If we could see such sequences in our collection, then we would know that the many unassigned sequences were also relevant and coded for proteins that had yet to be discovered.

The approach succeeded beyond our wildest expectations. We prepared a cDNA library from rabbit muscle and did shotgun sequencing of randomly isolated clones. Altogether about 20,000 nucleotides were manually determined (a large number in that era). Among the clones we identified were coding sequences for 13 different muscle proteins, in addition to a new iso-type of one protein (troponin T). This result gave us some confidence that, though some of the unassigned sequences might have arisen from artefacts, many of them undoubtedly encoded new proteins.

For many years we received requests for specific clones that people could use for their own purposes. (I had agreed with the editor of *Nature* before publication of the paper that we would freely distribute these clones.) Subsequently, the concept of ESTs and the strategy of shotgun sequencing were developed and expanded in the genome project^{2,3}, and included its commercial applications.

I have long been a proponent of and participant in the biotechnology industry, although not with respect to the work described here. It should not be forgotten, however, that the origins of success for many biotechnology enterprises are in basic research sponsored by a government or charitable foundation. I believe that this is as it should be.

But my fear is that the public at large may believe that private enterprise *per se* is sufficient to generate the discoveries and advances needed for the future. This is simply not true. Only the government and private charities have the capacity and vision to let an investigator pursue questions that come from his or her own natural curiosity. Eventually, the pursuit of pure curiosity bears fruit for society when new knowledge and concepts — such as ESTs and shotgun sequencing — are taken up by industry, where true applications can be financed.

Paul Schimmel

The Skaggs Institute for Chemical Biology, The

Scripps Research Institute, La Jolla, California 92037, USA

1. Putney, S. D., Herlihy, W. C. & Schimmel, P. *Nature* **302**, 718–721 (1983).
2. Adams, M. D. *et al. Science* **252**, 1651–1656 (1991).
3. Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C. & Venter, J. C. *Nature Genetics* **4**, 373–380 (1993).

When Brazilians achieve, it's against all the odds

Sir — As a young Brazilian scientist, I am proud to see a major achievement of the Brazilian scientific community in the pages of *Nature* — the sequencing of *Xylella fastidiosa* (*Nature* **406**, 151–157; 2000). However, I take issue with the view (*Nature* **406**, 109; 2000) that this milestone is "a signal to Brazil's young scientists that they do not need to leave the country to engage in world-class science". In fact, the achievement is not due to, but despite, Brazil's support for basic scientific research.

Most of Brazil's scientific productivity comes from the country's richest state, São Paulo, which supported the consortium that sequenced *X. fastidiosa*. The disparity between research support in São Paulo and in other Brazilian states is enormous: this research effort in no way represents the overall state of Brazilian science. Yet even in São Paulo scientists are grossly underpaid, often waiting years for salary readjustments that failed to keep up with high inflation rates in the 1980s and early 1990s. Few Brazilian students or lecturers will have a chance to read this or any other recent issue of *Nature*, as severe funding cuts have resulted in the cancellation of periodical subscriptions in many universities.

The rigid bureaucracy of Brazilian public universities turns the simplest transaction into a nightmare. On a recent visit to a top university, I had to make several trips between buildings to obtain authorizations to get a simple chemical from a stockroom. Young lecturers have to make huge efforts to assemble a laboratory.

Given the precarious situation of basic research in Brazil, it would be a serious mistake to let one major achievement project a view of a scientific landscape that is far from reality. Major changes in the administrative structure of Brazilian universities are needed to make them compatible with the development of fledgling scientific careers.

For most Brazilian scientists with academic positions in the United States or Europe, returning home remains akin to academic suicide. Reversing this situation may be far more difficult than sequencing and assembling whole genomes.

Tomas A. Prolla

University of Wisconsin–Madison, Departments of Genetics and Medical Genetics, 445 Henry Mall, Madison, Wisconsin 53706, USA

The gift of Lepidoptera

A literary giant and also a great butterfly systematist.

Nabokov's Butterflies: Unpublished and Uncollected Writings

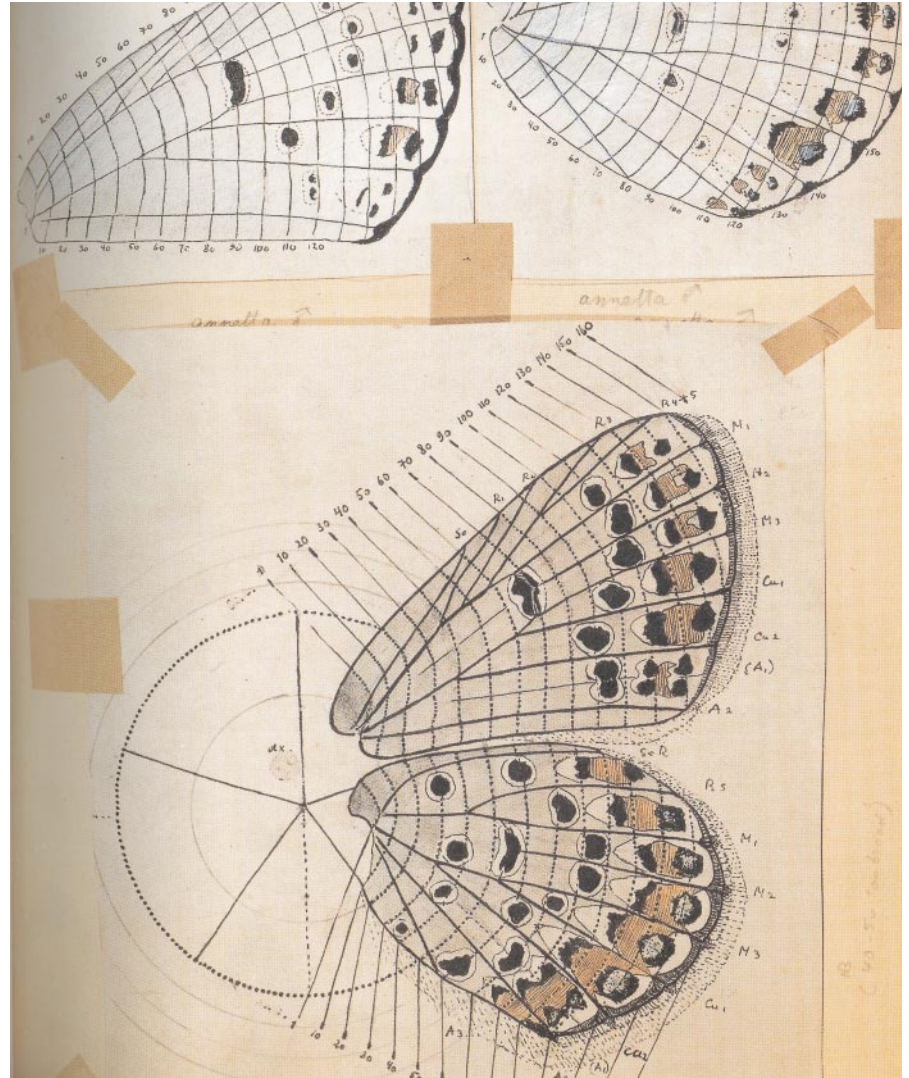
edited by Brian Boyd & Robert Michael Pyle
Allen Lane/Beacon: 2000. 800 pp. £25/\$45

James Mallet

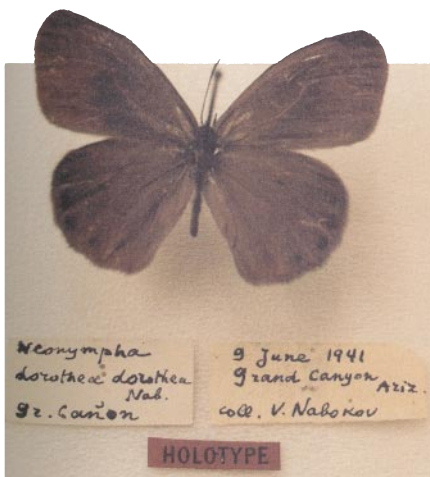
Butterfly collecting is often seen as a perversion — look no further than John Fowles' *The Collector* for an example. In *Nabokov's Butterflies*, which is a collection of extracts from Vladimir Nabokov's novels, poems, letters, interviews and scientific papers on butterflies, Brian Boyd and Robert Pyle have done a good job of dispelling that myth with regard to Nabokov. Yet an obsession with butterflies was indeed central to Nabokov's life. His knowledge of the Lepidoptera both distinguishes him from all other novelists and contributes strongly to the impact of his literary work. Butterflies are in every Nabokov book, and his novels were often written while collecting. "Butterflies help me in my writing. Very often when I go [on a collecting trip], and there are no butterflies, I am thinking. I wrote most of *Lolita* in this way."

Nabokov's novels are rich in details about natural history, stemming from the acute observations that enabled him also to become a great butterfly systematist. His accurate descriptions of plants and insects make a refreshing change from the tosh about nature found in many novels. The editors explain that the main purpose of *Nabokov's Butterflies* is to describe the connection between butterflies and Nabokov's art. Their other aim is to introduce Nabokov's writing to butterfly enthusiasts such as myself. They succeed on both counts.

The longest piece in the book is *Father's Butterflies*. Although intended by Nabokov



Spot on: wing-mapping enabled Nabokov to identify fine distinctions between races of 'blues'.



Winged legacy: Nabokov's wood nymph, finally named *Cyllopsis pyracmon nabokovi*.

as a postscript to his novel *The Gift* (1937), it was never included in any of the book's editions. The rise of Hitler led to this extraordinary manuscript fragment lying unread until it was translated by Nabokov's son Dmitri especially for this collection.

Many believe *The Gift* itself to be Nabokov's greatest novel. It is also the last novel that he wrote in Russian. The book is highly erudite, and teems with allusions to nineteenth-century Russian literature (it is this literature that is the 'gift' of the title). Its prose is interwoven with hidden poetry, as well as hidden meaning. Written in Berlin for the Russian émigré literary circle, its multi-layered story tells of an aristocratic poet and author, Fyodor Godunov-Cherdyntsev, who lives in Berlin. Fyodor writes a biography about a nineteenth-century Russian writer, and this biography itself forms a chapter in

The Gift. The book's publication causes a furore — Fyodor dares to mock the Russian literary giants of the previous century (as Nabokov does, of course, in *The Gift*). The book ends as Fyodor prepares to write another book (*The Gift* itself, according to Nabokov's foreword).

Father's Butterflies explains how Fyodor acquired a fascination for butterflies from his father. Fyodor's father is the Asian explorer and butterfly collector that Nabokov himself had wished to become before the revolution. Lepidopterists will identify immediately with Fyodor's boyhood symptoms of butterfly mania. Fyodor's affliction was contracted from viewing the rows of neatly pinned specimens in the glass-topped drawers of his father's collection; also from butterfly books and journal articles, by rearing caterpillars and collecting the imagoes in the woodlands

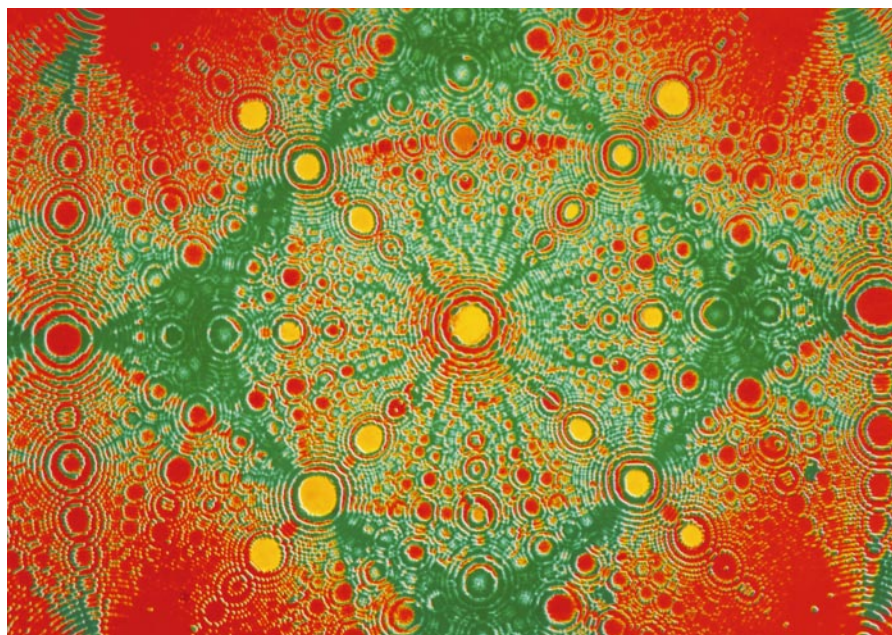
and meadows of his native Russia. Nabokov saw his own love of butterflies as a disease that his father before him had contracted from one of his German tutors: "from one of the latter, [my father] caught and passed on to me the *passio et morbus aureliani*."

According to my colleague Igor Emelianov, butterfly collecting was introduced into Russia from Germany by Catherine the Great, and remained a hobby of the aristocratic élite. Similarly, literature such as *The Gift* is hardly for the masses. With its combined textual and lepidopterous obscurities, *The Gift* becomes almost a symbol of the sophistication, decadence and weakness that led to the Bolshevik revolution. This is not to belittle Nabokov; the flaw of élitism demonstrated by *The Gift* is perhaps as threatening to our own technocracy as it was to the Russian aristocracy of 80 years ago.

Still, without the butterflies and the crystalline natural history, Nabokov's novels and poems, especially in their later and lighter English style, would have lacked much of their power. But how good a scientist was he? Nabokov was no dabbler. He discovered new butterfly taxa (he was inordinately proud of 'Nabokov's wood nymph' — *Cyllopsis pyracmon nabokovi*) and wrote systematic treatises in reputable entomology journals. His knowledge of Eurasian and New World butterflies was encyclopaedic and respected. He variously contemplated both a book of the butterflies of Europe and one on the butterflies of North America long before the standard works by Lionel Higgins and Norman Riley (*A Field Guide to the Butterflies of Britain and Europe*, 1970) and Alexander B. Klots (*A Field Guide to the Butterflies of North America*, 1951) were conceived.

Nabokov held some curious views. He hated the name 'red admiral', which his knowledge of early entomological literature enabled him to see was a corruption of the more appealing 'red admirable'. He believed that mimicry between poisonous and non-poisonous species was too exact to be explained by natural selection, and his systematic papers, with their elaborate and peculiar analyses of spot patterns on the wings of 'blues' (Lycaenidae), were sometimes derided as incomprehensible by other lepidopterists. But in the main, his systematic revisions are still important today. According to the butterfly taxonomist Gerardo Lamas, "had he become a professional lepidopterist, he would have been outstanding, despite his rather odd ideas on mimicry and other evolutionary subjects".

What drove this unique novelist-lepidopterist? Nabokov's *Butterflies* gives clues in a previously unpublished lecture to Russian literature freshmen at Wellesley College: "Whichever subject you have chosen, you must realize that knowledge in it is limitless. And yet there is a semblance of consolation within this dismal state of affairs: in the same



Once the ideal of perfect symmetry is abandoned, a deeper understanding may follow.

way as the whole universe may be completely reciprocated in the structure of an atom, an intelligent and assiduous student may find a small replica of all knowledge in a subject he has chosen for his special research. And if, upon choosing your subject, you allow yourself to be lured into the shaded lanes that lead from the main road you have chosen to the lovely and little-known nooks of special knowledge, if you lovingly finger the links of the many chains that connect your subject to the past and the future, and if by luck you hit upon some scrap of knowledge referring to your subject that has not yet become common knowledge, then will you know the true felicity of the great adventure of learning."

To which I am sure they replied, "Will that be in the test, sir?" ■

James Mallet is at the Galton Laboratory, Department of Biology, University College London, 4 Stephenson Way, London NW1 2HE, UK.

The bondage of symmetry broken

In Our Own Image: Personal Symmetry in Discovery

by István Hargittai and Magdolna Hargittai
Kluwer Academic/Plenum: 2000. 235 pp.
£34.50, \$49.95

David Blow

Hermann Weyl's classic book *Symmetry* concentrated on the interplay between the mathematical operation of symmetry, the success of the artist and the satisfaction of the beholder. In delightful contrast, István and Magdolna Hargittai emphasize the deadness of perfect symmetry, citing Pierre Curie's

"dissymmetry makes the phenomenon", and Fedorov's "crystallization is death". They follow, in fascinating detail, lines of development where the relaxation of an imposed or imagined symmetry has led to the deepening of scientific insight.

Take for example Johannes Kepler, who tried to link the orbits of the six known planets to the five Archimedean solids: the tetrahedron, cube, octahedron, dodecahedron and icosahedron. He thought that a series of concentric spheres including the orbits of the planets must have some special geometric property. A cube inscribed in the sphere of Saturn's orbit would just contain the sphere for Jupiter. A tetrahedron within that sphere would contain a sphere for Mars. And so on, until Mercury's sphere lay neatly within an octahedron whose apices touched the sphere for Venus. The idea worked pretty well, but did not quite fit the distances Copernicus had calculated between the planets. In trying to eliminate the assumed error, Kepler observed that planetary orbits are not circular but elliptical. He discovered the rules that link these ellipses with the length of time a planet takes to orbit the Sun, rules that Isaac Newton was able to generalize into a law of gravitation. Ideal symmetry was abandoned, leading directly to more profound understanding.

Or consider Jean-Baptiste Biot's discovery in the nineteenth century of optical rotation by organic solutions, exploited by Louis Pasteur. Suggestions of tetrahedral valencies for the carbon atom led Jacobus van't Hoff and Achille le Bel to recognize the possibility of stereoisomers, and chiral chemistry was born. Georges Friedel found that X-ray diffraction properties were centrosymmetric, but Coster, Knol and Prins showed this was not exactly true. Bijvoet exploited these

departures from symmetry and used them to confirm the 'Fischer convention' regarding configuration at an asymmetric carbon atom. Under the leadership of Vladimir Prelog and John Cornforth, a complete stereochemistry was built up: a system of classifying molecules as left-handed or right-handed, allowing us to explain the specific way in which enzymes interact with biological molecules and catalyse reactions.

There is always a wonderful interplay between scientific and artistic endeavours. While Buckminster Fuller was learning how to build geodesic domes, in which sheets of equilateral triangles were made to curl up by introducing occasional vertices where only five triangles meet, Aaron Klug and Don Caspar were studying the symmetry of spherical viruses. They found that viruses use just the same principles to build their capsids. But it took almost 30 years to discover that carbon can do the same trick on its own, in the C_{60} form that Harold Kroto named buckminsterfullerene.

Meanwhile, despite Leonhard Euler's proof that five-fold lattice symmetry could not exist, Roger Penrose showed how a flat surface could be covered with tiles based on the pentagon angles 72° and 144° . Alan Mackay found that such tilings, although lacking complete symmetry, create a diffraction pattern with five-fold symmetry. This ultimately led the International Union of Crystallography to refine its definition of a crystal.

These are just a few morsels from the lavish feast offered by *In our Own Image*. Another tells how the symmetry between matter and antimatter, envisaged by Paul Dirac, had to be relaxed when Chin-Shiun Wu's experiment demonstrated the lack of mirror symmetry in the decay of ^{60}Co gamma radiation in 1956. Curie's observation that lack of symmetry creates a phenomenon came into its own, and in the work of Abdus Salam, Steven Weinberg, Leon Lederman and others this relaxation allowed the rules of modern physics to be rewritten.

The story is fascinatingly instructive, and displays scintillating gems of scientific thought. Written by avid collectors of oral history, the book includes lengthy quotations from interviews with contemporary scientists, some evidently published here for the first time. These quotations bring the story to life, and generate insights that could not be gained from reading the scientific literature.

The book is not, however, an historical work. It lacks the completeness and thoroughness that a professional historian should provide. Its contents are to some extent dictated by the authors' idiosyncrasies. I was thrilled by unfamiliar quotations from Lucretius and Thomas Mann, but surprised at no mention of D'Arcy Thomp-

son's *On Growth and Form* (Cambridge University Press, 1917). In discussion of the handedness of the Universe, it seems a pity to pass over S. F. Mason's assertion that, because of the electroweak forces, D-sugars and L-amino acids are more stable than their enantiomorphs, or mirror-image forms.

Each chapter is dedicated to an iconic figure of science and art, and contains a series of linked short essays. Like the faces of a polyhedron, the subject matter of one essay fits closely to several others, and the authors have personally chosen a structure that is fascinatingly non-symmetrical. ■

David Blow is at the Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK.

Expanding horizons

The Accelerating Universe: Infinite Expansion, the Cosmological Constant, and the Beauty of the Cosmos

by Mario Livio

Wiley: 2000. 274 pp. \$27.95, £18.50

Francesco Bertola

The 18 December 1998 issue of *Science* awarded its accolade of "Breakthrough of the Year" to an astronomical discovery. The magazine's definition of a breakthrough is an event that profoundly changes the practice or interpretation of science. Indeed, if this astronomical discovery is completely and fully verified, it will constitute one of the major advances in cosmology of the last decades of the twentieth century.

A few years ago, two teams of astron-

omers — the Supernova Cosmology Project led by Saul Perlmutter of the Lawrence Berkeley National Laboratory, and the High-Z Supernova Search Team led by Brian Schmidt of Mount Stromlo and Siding Springs Observatories — undertook an observational project to generate the Hubble diagram, a plot of redshift against inferred distance, for distant galaxies with ages roughly half that of the Universe. They used type Ia supernovae as standard candles — whose distance can be calculated by comparing their brightness with an assumed absolute luminosity — taking advantage of the fact that supernovae are highly luminous and can be detected at enormous distances from the Earth.

The results of this work, presented in 1998, indicate that our Universe will expand for ever. In fact, the Hubble diagram for type Ia supernovae clearly suggests that the Universe's expansion proceeds at increasingly higher speeds. In other words, the Universe is accelerating. The long-debated question about the future of the Universe, whether the expansion will slow and end in a collapse or whether it will continue for ever, may have finally been answered. The consequences of this discovery are enormous, suggesting that our Universe is dominated by a 'dark' energy characterized by negative pressure that causes the acceleration. It marks a return to the cosmological constant, first introduced by Einstein, but always distasteful to the developers of cosmological models.

The title of Mario Livio's book points directly to this new discovery about the Universe. He provides a detailed account of the facts that led to this extraordinary result, presenting them within the broader framework of modern cosmology, and explaining even complex concepts clearly. Of course, as head of the Science Division at the Space Telescope



Cosmological constant: our obsession with the ways of the Universe remains unchanging.

HERMITAGE, ST PETERSBURG, RUSSIA/BRIDGEMAN ART LIBRARY

Science Institute in Baltimore, Livio has the great advantage of having been in close contact with the protagonists of the research he is describing, even if not a protagonist himself, and the reader feels this immediately. And, as a whole, the book brings the latest developments of present-day cosmology to the general reader.

The expansion of the Universe and its geometry, the presence of dark matter, the latest developments in the theory of inflation and the problem of origins are the main topics discussed by Livio. An extended chapter on the origin of life introduces us to his ideas about the anthropic principle, which considers our Universe fine-tuned for life. Livio finds the anthropic principle unconvincing for a number of reasons, among them the lack of predictive power.

He also discusses the methodology of the scientific process and defines a good and valid theory of the Universe. According to Livio, a theory must be beautiful, and he tries to explain this by saying that a theory has to fulfil the requirements of symmetry and simplicity. It must also obey the generalized Copernican principle that the best theories do not need us, as observers, to occupy a privileged position in the Universe in order to be valid. Although this concept about the development of scientific thought is a personal one, it is stimulating to follow Livio's description of how his ideas may or may not be applicable. The introduction of the cosmological constant, for example, could lead to the disappearance of beauty, but only "apparently", according to Livio.

The third theme of the book is the relationship between art and science. As an art enthusiast, Livio is in the position to make comparisons between the scientific process and that which leads to the creation of a work of art. His fascination reflects a wide public interest in this area in recent years. For example, the Venice Conferences on Cosmology and Philosophy, devoted to the themes of "The Beauty of the Universe" and "Art and Science", and the series of meetings on "The Inspiration of Astronomical Phenomena", provide the opportunity for astronomers to engage in direct discussion with artists.

And then there are the great art exhibitions — such as "Cosmos", currently running in Venice at the Palazzo Grassi (following stints in Montreal and Barcelona) — which can be considered as the commentary of artists on the great achievements of cosmology in the past two centuries.

There is only one disappointment with the book: the lack of reproductions of the numerous paintings mentioned in the text. Perhaps this is the book to carry when visiting the appropriate galleries. ■

Francesco Bertola is in the Department of Astronomy, University of Padua, Vicolo dell'Osservatorio 5, I-35122 Padua, Italy.

Science in culture

Angles on angels

A millennial exhibition of contemporary art at the Castel Sant'Angelo in Rome

Martin Kemp

In this tumultuous year of the Jubilee, Rome inevitably moves forwards by looking backwards. Venerable buildings have been rejuvenated, ancient marbles are restored to view, and once-sleeping museums rise to reassert their cultural richness. In this predominantly retrospective context, it is striking that a structure deeply redolent of the Classical and Christian pasts of the Holy City should be hosting a subtly subversive exhibition of experimental art that (among other things) places new technologies in the forefront of image-making.

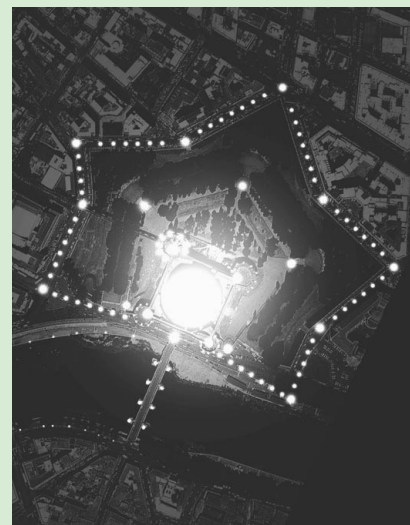
The Castel Sant'Angelo, assertively militant in its squat and bulbous mass, possesses an extraordinary history, embodying the essence of Rome. Gigantically constructed as Hadrian's imperial mausoleum, it was spectacularly recast and adorned by successive popes to serve the ballistic science of holy war — with bastions incongruously named after the Evangelists.

Within its stony bowels and sunlit terraces, it now houses "L'Assenza Inadente del Divino" ("The Pervasive Absence of the Divine"), curated by Franco Speroni and Luisa Valeriani, with site-specific works by Haim Steinbach, Joseph Kosuth, Studio Azzurro (a Milan-based group), Grazia Toderi and Ciriaco Campus.

The premise of the show is the presence (or otherwise) of the unknowable, the ineffable, the divine, in an age of scientific certainties and technological mastery. The core resides in a dialectic between the statement attributed to Saint Augustine that "God is better known by not knowing him" and Nietzsche's iconoclastic declaration that "God is dead".

The resolution lies in understanding how and what we are able to see from our constrained perspectives, and what we cannot see or even visualize through our outer and inner faculties of vision. The father-figures of the exhibition are those great inverters of conventional viewpoints, John Cage, one of whose anarchic musical compositions plays in a vaulted space enclosing a large grain mill, and Marcel Duchamp, who 'lectures' us in a cavernous chamber housing a mighty machine for grinding cannon shot.

High on the bright terraces, two coupled works, *La Via Mistica della Tecnologia*, by the 37-year-old Italian artist Grazia Toderi speak eloquently of the issue of perceptual and conceptual perspectives, earthly and heavenly, in complex interplay with the Castello itself. The tone is set by her linear installation of blue



Grazia Toderi's *La Via Mistica della Tecnologia*, rotating aerial video projection of the Castel Sant'Angelo and its environs.

lights demarcating the perimeter of the castellations, as if denoting a landing zone at an airport.

The implications are elaborated in a video of a rotating aerial view of the Castello and its environs, subtly manipulated by computer. In the former, we are invited to look upwards, from Earth to heaven, while in the other we stare downwards, courtesy of scientific technologies, from the former perspective of God and his attendant angels. An apparent 'runway' of lights is formed by the illuminations on the bridge over the Tiber. The great bronze angel looming over us at the summit of the fortifications implicitly lands and takes off from the papal aerodrome under the guidance of high-tech control systems.

It is all a question of points of view. From below we see the order of the heavens; from the elevated viewpoint now granted to us by our technologies, we directly see (angel-like) the stellate plan and hidden orders of the city as organism. Toderi's insight is eternally valid, whether in an age of faith or an age of reason. We see what we want to see and can see from our particular viewpoint. The hardest of all perceptual and conceptual tasks is to see the whole picture from another visual and mental place — which is what religions have required of us over the ages. ■

Martin Kemp is the Department of the History of Art, University of Oxford, 59 George Street, Oxford OX1 2BE, UK.

"L'Assenza Inadente del Divino" remains at the Castel Sant'Angelo until 31 December.

The man who knew doses

John Gaddum did much to define what it means to be a pharmacologist.

Rod Flower

When Sir John Gaddum, born 100 years ago, began his career in pharmacology it was a fledgling discipline, dating back only to the beginning of the twentieth century. Gaddum played a vital part in establishing the field, which he described as “a mingling of materia medica with physiology”, as a separate entity with its own viewpoint and set of questions. Of his many seminal research contributions, his work on the standardization of drugs and the theory of competitive drug antagonism were probably the most influential. His labs in London, Edinburgh and Cambridge were of international renown and played host to some of the finest experimentalists of the century.

In the 1950s, Gaddum’s address “The Science of Pharmacology” to the American Society for Pharmacology and Experimental Therapeutics was a manifesto for the emergent discipline and an agenda for future research. Among his top priorities were “finding out how drugs work” and expunging remedies of dubious provenance and efficacy from the bloated pharmacopoeias.

Gaddum was an accomplished mathematician who believed that no scientific discipline had come of age until it was quantitative. The statistical techniques he introduced provided the tools for pharmacology, and he once joked that materia medica was the only field of knowledge “that has become smaller as it has advanced”. Other tasks set by Gaddum for his colleagues included the study of drug toxicity and pharmacokinetics, and the participation with clinicians in the design of clinical trials (for which he also laid much of the theoretical foundations). Discovering new drugs was, of course, also on Gaddum’s agenda, but to him this was a secondary goal. This may seem odd, but perhaps it reflects his early interest in sailing, which taught him that one must know one’s position to plot an accurate course to one’s destination.

Has pharmacology lived up to Gaddum’s expectations, and is his agenda still valid? In the 1930s, Gaddum briefly occupied the chair of pharmacology at Cairo University and it is salutary to remind ourselves how few really effective drugs there were at that time. His wife, Iris, who worked there as a dermatologist, later recalled that only morphine, aspirin and coal tar were available in her clinic. Apart from making a diagnosis, there was little that she could do for her patients except to ensure adequate nursing care. Topical steroids later revolutionized dermatology and many other



Jack of all trades: Gaddum realized that many disciplines are needed to discover and understand drugs.

therapeutic triumphs have since fundamentally changed other branches of medicine. Even Gaddum later remarked that pharmacologists had been “almost too successful”. Ironically, those very pharmacopoeias that he and his contemporaries strove to prune have burgeoned in size once again.

There has been good progress over the past 50 years with another item on Gaddum’s agenda — discovering how existing drugs actually work. Aspirin and morphine, for example, have yielded up many, if not all, of their secrets. But as the 1998 Nobel Prize in Physiology or Medicine shows, there is still much to learn about even such apparently simple drugs as nitrates.

The fundamental problems of pharmacology remain as Gaddum defined them half a century ago, although the intellectual landscape is very different. Gaddum’s early research training was in physiology at a time when, as a colleague put it, all he needed to do to become a pharmacologist was “to learn doses”. He later did much to define exactly what it means to be a pharmacologist; his 1940 textbook *Pharmacology* ran to five editions in several languages.

Gaddum always maintained that phar-

macology is truly interdisciplinary by its very nature, describing the pharmacologist as a “jack of all trades”. To ask how a drug works, what changes it produces and how it can be improved is to invite the assistance and insights of specialists from many other disciplines. This is truer today than ever, for of course the future of pharmacology is inextricably linked with that quintessentially millennial activity, the Human Genome Project, and will depend on many new professional alliances.

Many medical schools have sought to encourage this type of research by abolishing the traditional departmental structures, but Gaddum was concerned about organized ‘team’ science. His concept of collaboration was a synergistic interaction between specialists, and he might have argued that this could not be achieved simply by changing the names on laboratory doors, but rather by building upon the unique contributions that each separate discipline can make to the business of drug discovery and to an understanding of drug action. ■

Rod Flower is at the William Harvey Research Institute (QMW), Charterhouse Square, London EC1M 6BQ, UK.

Words, words, words

Kissing a biotechnological blarney stone

Elisabeth Malartre

Richard calls Marilyn's normal morning routine "the old gal doing her drugs". She takes antioxidants to keep young; a hormone-replacement pill; a diuretic for high blood pressure; vitamins ("just because"); and a soft gel called Palaver, all with calcium-enriched orange juice. Then she has breakfast and reads the paper, working on the crossword between trips to the bathroom. She rarely answers the phone or goes out before 10:00 a.m., letting her pills do their work.

A bit tedious, perhaps, but she doesn't really mind. Palaver makes it all worthwhile.

"I still don't see why you have to hog the crossword every morning," he grumbles.

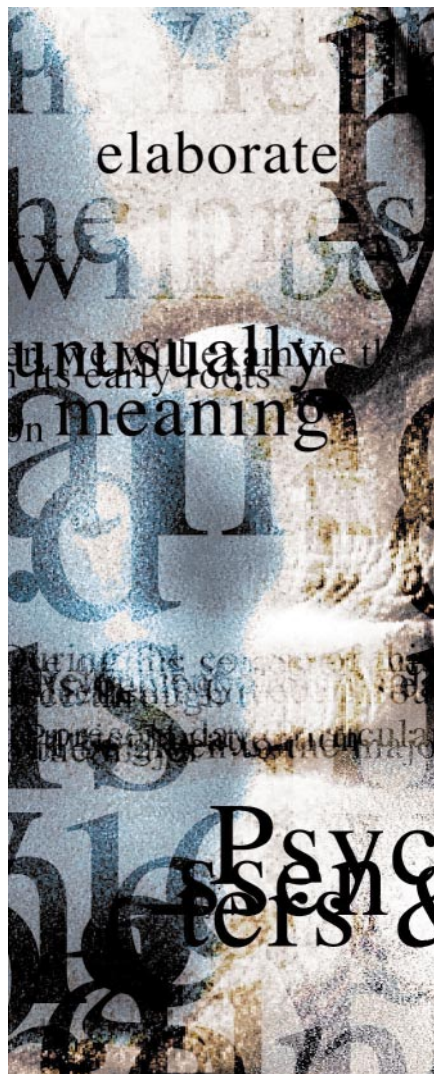
She smiles: "It works like your Viagra, Richard, dear. You have to stimulate the brain to get it to work right."

It was a fortuitous discovery. A graduate student named Anne Cashmore, trying to settle once and for all whether chimpanzees had language capability, was looking at brain organization in detail. While closely examining MRI slices of Broca's area in humans, she discovered a tiny sac-like organ, filled with dense material. Under the scanning electron microscope, it was found to be a heterogeneous population of discrete rod-like organometallic particles, up to a few microns long. Thin EM sections hinted at definite but non-regular internal structure. Only at the highest magnification did their true nature manifest themselves: they were words.

Cashmore wrote up her results, suggesting that the sac functioned like a gall bladder, storing words that were then distributed in a still-mysterious process to other neurons. Every journal turned her paper down. Then someone leaked it to a *New York Times* science stringer. After that, it was headline news. Websites blossomed with her micrographs. Talk shows were in a frenzy. Cashmore's discovery was quickly dubbed the 'word sac' by the media.

The brain-research community immediately denounced it as a hoax: it violated every known precept about brain organization and function. "Chomsky *ad absurdum*," spouted the dean of the field. Cashmore was almost thrown out of graduate school.

But even her staunchest critics were silenced as corroborative results appeared: the size and number of the particles correlated with the age of the person, up to young adulthood. Politicians and others who talk for a living were found to have unusually



large sacs, but the sacs of autistic children were empty. There was also a small but statistically significant correlation between the size of the particles and level of educational attainment. People "speaking in tongues" turned out to have gibberish words in their sacs, accessed by emotional storms in the cerebrum. As in the case of the Shroud of Turin, however, true believers rejected this explanation. Finally, Cashmore found that chimpanzees' sacs were quite small, with particles roughly the size of those in three-year-old humans. That resonated with the observations made by early researchers on chimpanzee language, and Cashmore was allowed to complete her thesis.

Even more interesting, perhaps, were the therapeutic ramifications. In an ageing population increasingly at a loss for words, there was suddenly hope. The then-infant Rose

Genomics, Inc. contracted with Cashmore's university to investigate the nature of the particles and develop any drug uses. The first crude attempt at a drug for perimenopausal aphasia was basically a slurry of particle material. Ingested or injected, it migrated to Broca's area, the way iodine concentrates in the thyroid. Within a few hours, new particles began to appear in the word sac, more in an active brain.

Those early researchers at Rose Genomics drank purified particles from brain donors, despite the terrible risks of catching viruses and prions. But it worked: the particles migrated to the word sac and were there for retrieval after an hour. Michael Rose himself, then in his fifties, was one of the experimenters. In his testimony before Congress, he said it worked "like packed red blood cells carrying extra oxygen for an athlete. The new words filled in the blanks for me when I talked." Then, at last, a German researcher at Rose Genomics discovered, in a still secret process, how to produce word particles synthetically, and the dam was broken.

The new drug was named Palaver™, and it was a huge instant success, eclipsing even Viagra. Both sexes take Palaver, and most people want it every day. Commercials splashed across screens everywhere trumpeted, "Enhance your speech — speak like a Professor" or, "Vitamins for your vocabulary".

Some of the early claims were not fulfilled. It isn't possible to ascertain a language simply by swallowing the words: you have to comprehend the grammar and sentence structure, then the sac supplies the words. Experiments with Chinese particles resulted in near-psychotic episodes in native English speakers. One man reported "a meaningless shower of sounds whenever I opened my mouth".

What you can do, however, is concoct designer vocabularies; or add a few grandiose words to your customary parlance. PalaverPlus™ can temporarily enhance a mundane vocabulary, "like a built-in thesaurus", although substitution mayhap leads to malapropisms or inappropriate selections.

So, every morning Marilyn takes her soft gel and waits for an hour, doing a crossword to warm up, priming the pump, so to speak. Meantime, the particles race to her brain, filling up the sac with all those lovely, formerly elusive, quanta of meaning. ■

Elisabeth Malartre is an environmental consultant and science writer, based in Orange County, California.

JACEY

Vesicle fiesta at the synapse

Gary Matthews

Communication between nerve cells is what makes our brains work. Several new studies offer us a close-up view of the chemical signals controlling these cellular conversations.

Nerve cells need to transfer information to each other, and researchers have been unravelling the secrets of how they do this for decades. Technical advances and human ingenuity have now combined to take us several steps further in this intellectual challenge. With the publication of two reports in this issue^{1,2} (pages 849 and 889) and a third in *Science*³, we can actually watch the intricate molecular dance that takes place when neurons talk to each other, and we learn more about the control mechanisms involved.

Neurons communicate at special junctions, known as synapses, where the transmitting cell releases a chemical signal into the small gap separating it from the receiving cell. When the transmitting neuron is stimulated, channels in its plasma membrane at the synapse open, allowing calcium ions to flood into the cell. This prompts sacks containing chemical neurotransmitters to fuse with the plasma membrane, releasing their contents — the 'signal' — into the synaptic gap. These neurotransmitters then diffuse across the gap to the neighbouring neuron, where they bind to receptors on the plasma membrane and trigger an electrical response. Zenisek *et al.*¹ have filmed the cellular sacks, known as synaptic vesicles, as they move to the membrane and fuse there in response to an influx of Ca^{2+} . Meanwhile, Schneggenburger and Neher² and Bollmann *et al.*³ have found that the sensor that triggers this fusion is more sensitive to Ca^{2+} levels than was previously supposed.

Physiological studies indicate that synaptic vesicles prepare for rapid, Ca^{2+} -triggered fusion with the plasma membrane (exocytosis) through a sequence of docking and 'priming' stages (Fig. 1). These preparatory stages and exocytosis itself have been directly observed in other types of cell that secrete hormones⁴. But the small sizes of synaptic terminals (with diameters of about 1 μm) and synaptic vesicles (diameter about 30 nm) have made it difficult to observe the events preceding neurotransmitter release.

Zenisek *et al.*¹ overcame these problems by using total-internal-reflection microscopy to study the giant (roughly 10 μm) synaptic terminals of goldfish retinal bipolar neurons. Their technique exploits the way a neuron's synaptic terminal will stick tightly to a glass coverslip. Light reflected at the interface between the glass and the terminal

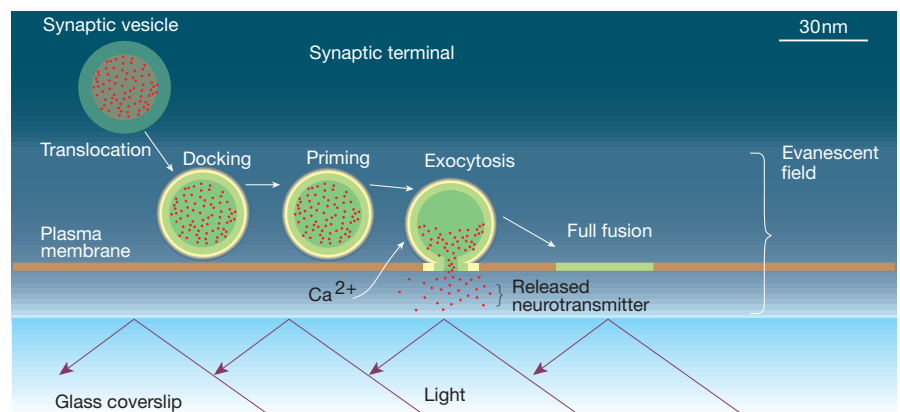


Figure 1 Watching neurotransmitter release. Communication between nerve cells takes place across the synaptic gap between them. Neurotransmitter chemicals are sent from the terminal of the signalling neuron across the gap to its neighbour. When a neuron is stimulated, calcium channels in the plasma membrane of the nerve terminal open, allowing an influx of calcium ions. This causes vesicles loaded with neurotransmitter (red) to move to the plasma membrane, dock, and undergo 'priming' for fusion (exocytosis). Zenisek *et al.*¹ used total-internal-reflection microscopy to watch the dynamics of Ca^{2+} -driven fusion of single synaptic vesicles. Light reflects from the interface between a glass coverslip and the neuron, establishing a region of brightness (an evanescent field) that extends a short distance into the nerve terminal. The vesicles are labelled with a fluorescent marker, and begin to glow when they enter the evanescent field. During exocytosis, the fluorescence spreads into the plasma membrane of the synaptic terminal, allowing exocytosis to be seen.

forms an evanescent field that excites any fluorescent molecules that come within about 100 nm of the interface⁵ (Fig. 1). When a fluorescently labelled vesicle enters this field, its brightness increases as it moves nearer to the plasma membrane. And when fusion takes place, the fluorescence is incorporated into the membrane. By labelling only a small fraction of the vesicles in the terminal, Zenisek *et al.* were able to see single vesicles approaching, docking and fusing with the membrane during stimulation of the neuron.

Their work also reveals several new features of vesicle dynamics. Fusion occurred preferentially at specific 'hot spots' along the membrane. These probably correspond to active zones, where Ca^{2+} channels cluster and most neurotransmitter release happens. During an electrical stimulus, vesicles that were already near the plasma membrane fused rapidly with it. But vesicles arriving at the membrane a little later paused, on average, for a quarter of a second before approaching the membrane and fusing. This pause might represent the time needed for priming, during which interactions between proteins on the plasma membrane and the vesicle's own membrane (Fig. 2, overleaf)

prepare the vesicle for exocytosis. During the pause, newly arrived vesicles often stopped about 20 nm away from the cell membrane. Active zones in retinal bipolar neurons are marked by synaptic 'ribbons', which tether numerous synaptic vesicles, and it may be these ribbons that hold the vesicles away from the membrane during the pause.

But there are some limitations to the microscopy technique. The closeness of the plasma membrane and the glass substrate might influence exocytosis. For example, as Ca^{2+} is taken up by the nerve terminal, the levels of Ca^{2+} available in the small space between the plasma membrane and the glass coverslip might drop, which could affect the kinetics of exocytosis for late-arriving vesicles. Also, Zenisek *et al.* detected fusion (and hence neurotransmitter release) by watching for fluorescence from the vesicle appearing in the plasma membrane, so only full fusion was detected. But neurotransmitters can sometimes be released from a vesicle that is not fully merged with the plasma membrane. Zenisek *et al.* saw some vesicles move to the membrane, pause, and then retreat without fusing. This behaviour might reflect the reversibility of docking, but it may also show neurotransmitter release without full fusion.

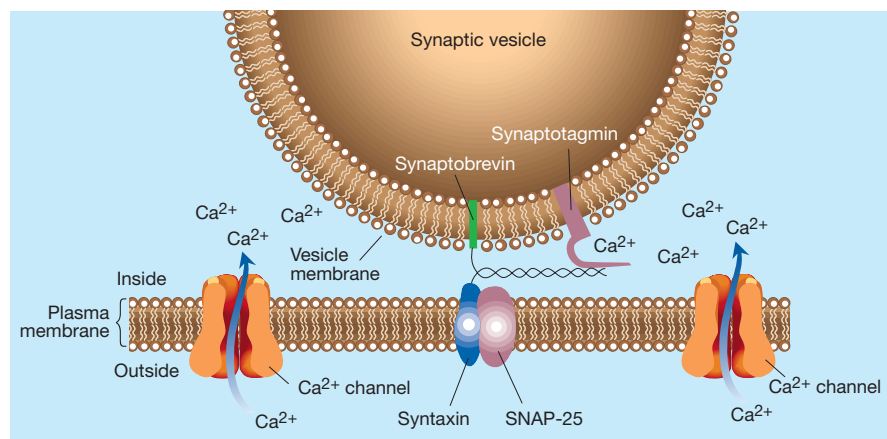


Figure 2 Kinetics of calcium-dependent neurotransmitter release. This close-up view of a nerve terminal shows a synaptic vesicle docked at the plasma membrane near two Ca^{2+} channels. During priming, the plasma-membrane proteins syntaxin and SNAP-25 and the vesicle protein synaptobrevin form a complex required for vesicle fusion with the plasma membrane⁹. The Ca^{2+} sensor that triggers fusion is thought to be the vesicle protein synaptotagmin⁹. Schneggenburger and Neher² and Bollmann *et al.*³ show that the Ca^{2+} concentration needed to trigger fusion is lower than previously thought. So the Ca^{2+} sensor does not need to be close to a single Ca^{2+} channel, where the Ca^{2+} concentration is very high. Instead it can be some distance away from the channels, where the Ca^{2+} concentration is lower as a result of diffusion. This means that more than one Ca^{2+} channel can influence neurotransmitter release. (Modified from ref. 10.)

Meanwhile, Schneggenburger and Neher² and Bollmann *et al.*³ have been looking at the way in which vesicle fusion depends on calcium. Averaged across the whole terminal, the concentration of Ca^{2+} inside the neuron rises from about 0.1 μM to just under 0.5 μM during neuronal stimulation. But vesicle fusion is thought to be triggered by the much higher Ca^{2+} concentrations (more than 100 μM) near the open mouth of individual Ca^{2+} channels⁶ (Fig. 2). It now appears^{2,3} that such high concentrations are not needed. Schneggenburger and Neher² and Bollmann *et al.*³ released 'caged' Ca^{2+} to produce known levels of Ca^{2+} in the calyx of Held — a large synaptic terminal that allows nerve impulses in both the signalling and the receiving neurons to be recorded simultaneously. The authors found that a rise in Ca^{2+} concentration in the signalling neuron to 10 μM or less was enough to drive fast neurotransmitter release. They estimate that an action potential (the electrical change in a neuron that follows stimulation) activates the exocytosis of synaptic vesicles by raising Ca^{2+} levels to 10–25 μM for about 0.5 ms.

Apparently, it is not important for the exocytosis machinery and Ca^{2+} channels to be near each other to allow rapid neurotransmitter release (Fig. 2) — Ca^{2+} can diffuse some distance away from a channel into the cell, its concentration declining as it diffuses, and still trigger neurotransmitter release. This means that single vesicles might be influenced by Ca^{2+} ions entering through several channels. If so, and if the Ca^{2+} sensor is therefore not saturated by concentrated Ca^{2+} influx through a single open channel, greater control of neurotransmitter release becomes possible.

These results^{2,3} differ significantly from those of similar experiments on retinal bipolar neurons⁷, where increases in Ca^{2+} concentration to 10–100 μM were needed to drive exocytosis. But unlike neurons in the calyx of Held, bipolar neurons release neurotransmitters over relatively long periods and do not fire action potentials, which tend to be brief. The global Ca^{2+} concentration can

exceed 1 μM in bipolar-cell synaptic terminals, even with moderate stimulation. So releasable vesicles would be exhausted rapidly if the exocytosis machinery here were as sensitive as that at the calyx of Held. Perhaps low Ca^{2+} affinity is a hallmark of sustained neurotransmitter release, whereas the brevity of Ca^{2+} signals driven by action potentials allows for higher Ca^{2+} affinity and greater integration of local Ca^{2+} signals.

It is almost 50 years since Katz and colleagues embarked on the work that still forms the cornerstone of our understanding of how neurons talk to each other⁸. But this field is as exciting as ever, and it is thrilling to be able to watch single vesicles as they fuse, to observe directly the stages before exocytosis, and to begin to glimpse the calcium signals that trigger fusion.

Gary Matthews is in the Department of Neurobiology and Behavior, State University of New York, Stony Brook, New York 11794-5230, USA.
e-mail: Gary.G.Matthews@sunysb.edu

1. Zenisek, D., Steyer, J. A. & Almers, W. *Nature* **406**, 849–854 (2000).
2. Schneggenburger, R. & Neher, E. *Nature* **406**, 889–893 (2000).
3. Bollmann, J. H., Sakmann, B. & Borst, J. G. G. *Science* **289**, 953–956 (2000).
4. Steyer, J. A., Horstmann, H. & Almers, W. *Nature* **388**, 474–478 (1997).
5. Steyer, J. A. & Almers, W. *Biophys. J.* **76**, 2262–2271 (2000).
6. Neher, E. *Neuron* **20**, 389–399 (1998).
7. Heidelberger, R., Heinemann, C., Neher, E. & Matthews, G. *Nature* **371**, 513–515 (1994).
8. Katz, B. *The Release of Neural Transmitter Substances* (Liverpool Univ. Press, Liverpool, 1969).
9. Südhof, T. C. *Nature* **375**, 645–653 (1995).
10. Matthews, G. *Neurobiology: Molecules, Cells and Systems* 2nd edn (Blackwell, Malden, MA, in the press).

Chemistry

Another noble gas conquered

Gernot Frenking

On page 874 of this issue Khriachtchev *et al.*¹ report the synthesis of the first compound containing the noble gas argon. This leaves only two stable elements in the periodic table — helium and neon — for which no neutral compound exists. But the experimental strategy that was successful for argon might also work for these elements.

One of the most fundamental questions in chemistry, which remained unanswered for a long time, was what drives the chemical elements to form stable molecules. Why are some elements very reactive, some less reactive, and a peculiar small group of elements, known as noble gases, not reactive at all? Quantum theory finally explained these observations by showing that the electrons within atoms are arranged in shells around the atomic nucleus. Each shell can only take a fixed number of electrons, and a completely filled shell is energetically very favourable. So the driving force behind chemical reactions



Figure 1 Argon is no longer alone. The first neutral argon molecule, synthesized by Khriachtchev *et al.*¹ is HArF, a stable hydride compound. The numbers give the calculated distances between the atoms in picometres.

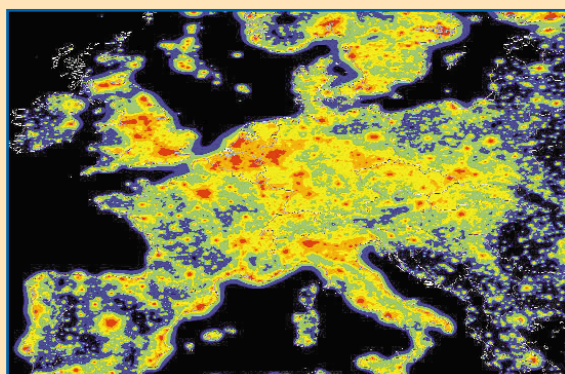
is the filling of electron shells by release or uptake of electrons, or by sharing electrons with neighbouring atoms, which leads to the formation of chemical bonds.

The noble gases have a completely filled outer electron shell, so there is no driving force for these elements to engage in chemical bonding, as the energetically most stable arrangement of electrons already exists in the neutral atoms. Many chemists believed that it was impossible to synthesize a stable molecule containing a noble gas element,

Astronomy

Lights go on all over Europe

Astronomers at last week's general assembly of the International Astronomical Union in Manchester expressed concerns about rampant light pollution. High-resolution satellite images reveal the extent of the problem (Cinzano, P. *et al. Mon. Not. R. Astron. Soc.*, in the press). This map of Europe shows the ratios between artificial and natural sky brightness (red is the worst at more than 9:1).



but others saw the reluctance of the elements helium, neon, argon, krypton, xenon and radioactive radon to form chemical bonds as a great challenge to experimental chemistry. They were supported in their quest by theoretical predictions made by the late Nobel prizewinner Linus Pauling³, who predicted in 1933 that molecules containing xenon and krypton should exist.

It took nearly 30 years before Pauling's theory was verified by experiment. The breakthrough came in 1962 when Neil Bartlett found that the strongly oxidizing agent platinum hexafluoride, PtF₆, is powerful enough to ionize (that is, to remove an electron from) an oxygen molecule, O₂, yielding the stable salt compound (O₂)⁺(PtF₆)⁻. Bartlett recognized that the energy needed to remove an electron from xenon is slightly less than the ionization energy of O₂, so he carried out the same experiment with xenon and PtF₆. The spontaneous reaction produced the first neutral compound to contain a noble gas³. Bartlett's discovery was soon followed by syntheses of many other compounds of xenon, and also of krypton, which has only a slightly higher ionization energy than xenon⁴.

This suggests that the ionization energy of the noble gases should be a measure of their willingness to form a chemical bond. The ionization energy, IE, is measured in electron volts, eV. Xenon (IE = 12.13 eV) compounds are indeed more stable than krypton (IE = 14.00 eV) compounds. Quantum chemical calculations predicted that neutral compounds of argon (IE = 15.76 eV) should also exist, because its ionization energy is still lower than that of fluorine (IE = 17.42 eV). Therefore, salt compounds of ArF⁺ with suitable anions might be stable, although synthesizing them would be a formidable task⁵. Compounds of helium (IE = 24.59 eV) and neon (IE = 21.56 eV) were not expected to be stable because of their very high ionization energies⁶.

Khriachtchev *et al.*¹ have now synthesized

the first stable neutral argon compound, HArF (Fig. 1). This is an important milestone in the history of chemistry, because the number of elements for which no neutral molecule has been synthesized is now only two. It is also an exciting accomplishment because the strategy used by the authors was different from the one suggested by theory. HArF is not a salt compound. More exciting is the fact that the type of noble-gas bonding in HArF might also be found in the neon compound HNeF, and even in the helium compound HHeF. In fact, theoretical calculations announced last year predicted that HHeF could be a stable neutral molecule⁷.

It should be noted that molecules such as HArF are only kinetically stable and exist only in a low-temperature solid matrix (in this case made from inert argon). The molecules rapidly decompose if they come into contact with each other or if they are warmed up, because they break up into unbound argon atoms and HF, which are much more stable species. It is only the energy barrier involved in breaking the weak H–ArF bond that prevents it from spontaneous decomposition. So HArF does not fall into the same category as the numerous xenon and krypton compounds, which can easily be prepared and handled in solutions or as solid salt compounds. Nonetheless, the work of Khriachtchev *et al.* is a further breakthrough in the field of noble-gas chemistry, and might even help chemists to conquer the two least reactive elements, neon and helium. ■

Gernot Frenking is at the Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Strasse, D-35039 Marburg, Germany. e-mail: frenking@chemie.uni-marburg.de

1. Khriachtchev, L., Pettersson, M., Runeberg, N., Lundell, J. & Räsänen, M. *Nature* **406**, 874–876 (2000).
2. Pauling, L. *J. Am. Chem. Soc.* **55**, 1895–1900 (1933).
3. Bartlett, N. *Proc. Chem. Soc.* 218 (1962).
4. Laszlo, P. & Schrobilgen, G. *Angew. Chem. Int. Edn Engl.* **27**, 479–489 (1988).
5. Frenking, G., Koch, W., Deakne, C. A., Liebman, J. & Bartlett, N. *J. Am. Chem. Soc.* **111**, 31–33 (1989).
6. Frenking, G. & Cremer, D. *Structure Bonding* **73**, 17–95 (1990).
7. Wong, M. W. *J. Am. Chem. Soc.* **122**, 6289–6290 (2000).



100 YEARS AGO

An interesting and detailed account of Count von Zeppelin's successful trial trip of his navigable balloon on July 2 is given in *Die Umschau* by an anonymous author, who has endeavoured to dispel the somewhat exaggerated reports which have been circulated as to the success or failure of the experiment. It is pointed out that the delay in the ascent, which some persons attributed to an accident, was really caused by the wind being too strong at the time originally proposed for the trip. The wind-velocity at the time of starting was 5.5 metres per second, and the balloon was actually driven forwards for a short time in the face of this wind... The wind causing the balloon to drift towards the shore, a descent was made in order that Count Zeppelin might land on the water (to use an Irishism), and thus have his machine towed back by steamer. The descent was very gradual, the cars gently sinking down to the water without the sudden jerk which is commonly experienced in an ordinary balloon.

From *Nature* 23 August 1900.

50 YEARS AGO

If we slide one solid body over another, there is a resistance to the motion which we call friction. What is the cause of this, and what is really happening at the interface between the solids during sliding? It is usually considered that friction is a nuisance, and from the earliest times man has made ingenious attempts to eliminate or to diminish it to as small a value as possible. It is not so much the work we have to do in overcoming the friction—the power loss—which is important, although this can be considerable. In a modern motor-car, for example, about 20 per cent of the power is wasted in overcoming friction... The real trouble is the damage that is done by the friction—the wear or seizure of some vital part of the machine. It is this factor, perhaps more than any other, which limits the design and which shortens the effective life of, say, an aero-engine or other complex machine. Nevertheless, the credit side is considerable: if the coefficient of friction were much less than it is, walking would be impossible and we should have to use cog wheels or perhaps some system of suction pads like an octopus to get about at all. Conversely, if it were much higher than it is, we should just stick fast wherever we were put.

From *Nature* 26 August 1950.

Motor proteins

Directing direction

R. A. Cross

Molecular motors are essential to the life of a cell, working as they do to transport cargoes of cellular components to their correct destinations. Motor proteins always move along linear tracks in a specific direction, which is dictated by something intrinsic to the protein itself. But how is this direction determined? This question is really a facet of the overall problem of how these motors move. It is important for understanding both their design principles and the more general issue of how motors organize cells.

A new and fascinating episode in the exploration of motor mechanism unfolds on page 913 of this issue¹, where Endow and Higuchi describe a set of mutations of single amino acids in a microtubule motor called *ncd*. Like other members of the kinesin motor family, this motor is powered by adenosine triphosphate (ATP) and carries its cargo along microtubule tracks (myosin motors, by contrast, move along actin filaments). When pinned to a surface, such as glass, these motors can move microtubules across the surface. The effects of the mutations in *ncd* described by Endow and Higuchi are, to say the least, bewildering. When coated onto glass, the mutant motors vigorously and sustainably shuttle microtubules first one way and then the other. The motors on the glass surface will even break the microtubules, and slide the two pieces off in opposite directions. The mind reels: how can the same motor molecule produce forceful progress in two opposite directions?

Most of the kinesin motors studied so far move towards the 'plus' ends of microtubules — the ends where subunits are added at a high rate. But some, including *ncd*, move in the opposite direction, towards the minus ends. All the kinesins have some sort of head-on-a-tail configuration (Fig. 1); the heads have the microtubule- and ATP-binding sites, whereas the tails bind to particular cargoes. To move, the motors grab onto microtubules with their heads and 'step' along. Some kinesins^{2,3} can take step after step, so a single molecule can walk its load along a microtubule on its own. The *ncd* motor cannot do this 'coordinated' walking, and probably relies instead on teamwork with other molecules to transport its cargoes.

Despite these differences, the heads of all kinesins studied so far are structurally and functionally similar. It is the tails — and specifically the way in which the tails are joined to the heads — that are different. For all known naturally plus-end-directed kinesins, the tail is attached to the carboxy-terminal

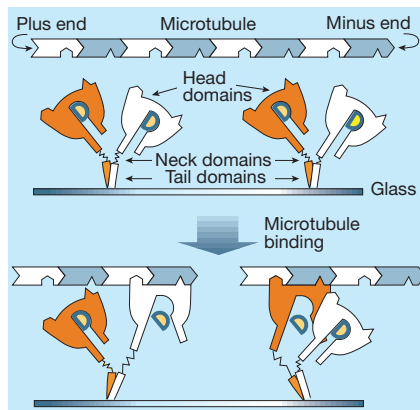


Figure 1 What determines the direction in which motor proteins move? Endow and Higuchi¹ have approached this problem by generating mutants of *ncd*, a motor that normally moves towards the minus end of microtubules. The mutants have single amino-acid changes in the neck or head domain. When attached to a glass surface, they can move microtubules in either direction, or can break them in two and slide the pieces in opposite directions. This diagram shows how they might do this. The mutations studied are proposed to alter the symmetrical relationship between the two heads of each *ncd* molecule, so the heads are poised in different attitudes towards the microtubule. Because of this, one head tends to find sites towards the plus end, while the other tends to find sites towards the minus end. Opposing forces can thereby develop, and might snap the microtubule, as shown.

end of the heads, whereas in all the naturally minus-end-directed kinesins, the tail is attached to the amino terminus. Swapping the tails around using protein engineering reverses the direction in which the motor moves, but only if the join is made in exactly the right place⁴. If the junction is made in the wrong place, or if a single mutation disrupts it, things stay roughly as they were. So something about the junction's structure, which presumably determines the attitude or range of movement of the heads relative to the first part of the tail, is crucial in setting the direction of stepping.

All well and good, if rather fuzzy, but the data from Endow and Higuchi¹ call even this simple idea into question. Somehow, in their mutants both plus- and minus-ended types of directionality occur in the same motor. How could this work? Perhaps wisely, Endow and Higuchi avoid detailed speculation as to what might be happening at the molecular level. But it does no harm to rehearse possible models.

The structures of the heads of plus-end-directed and minus-end-directed kinesins

are similar, and it is likely that the changes in conformation that these heads undergo as they move along microtubules are also similar. For example, the heads of both *ncd* and the prototypical plus-end-directed kinesin-family member, kinesin itself, bind to microtubules in a similar way⁵, with one binding site per tubulin heterodimer (the basic unit of microtubules). The heads of these two motors also undergo similar conformational changes as they move⁶. So, what we need is a model, or models, in which essentially similar conformational motions, associated with ATP hydrolysis by the heads, are converted by a kind of gearbox to drive motion in opposite directions.

One clue may lie in the symmetry of the new results — roughly half the microtubules move one way, half the other, at roughly equal speeds. This functional symmetry might originate in the physical symmetry of *ncd*⁷. In the wild-type *ncd* molecule the two heads point in exactly opposite directions. Only one head at a time can interact with microtubules^{8–11}, and this interaction always produces minus-end-directed force.

But suppose that, in the mutants, the symmetry is changed. If rotation about the bonds in the junction between the head and the tail is allowed¹², then the heads can become non-equivalent. One head might tend to diffuse to binding sites ahead in the minus-end direction, and the other to sites ahead in the plus-end direction (Fig. 1). In this kind of model, the direction of force is entirely a result of directional bias in the diffusional 'searching' process by which the heads find and bind to the next site along the microtubule track¹³. Subsequent conformational changes, possibly involving reversible interactions between the head and the first part of the tail (as happens in walking kinesins⁵), are directional only because the initial binding to the microtubule is directional. Otherwise, any movements contributed by conformational changes would be counterproductive for progress in one of the two directions.

One way to test this model would be to find out whether all of Endow and Higuchi's *ncd* molecules can step in both directions, or whether it is only the population of *ncd* motors that is bidirectional, with individual molecules being somehow committed to move only one way. Endow and Higuchi tackled this question by attaching single motor molecules to micrometre-sized beads, and bringing the bead-motor complexes into contact with a microtubule using a single-beam optical trap. They then tracked the motion of the beads. When they did this with normal *ncd*, the beads moved only towards the microtubule minus end. With the mutants, the kicks were in both directions — so individual *ncd* mutants can indeed move in either direction.

But how far, and in what direction, can single heads move? This is the key question

now, and might be best approached by using a dual-beam optical trap, which is better for detecting short-range movements and has already been used to detect single myosin movements¹⁴.

R. A. Cross is at the Molecular Motors Group, Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 0TL, UK.

e-mail: r.cross@mcri.ac.uk

1. Endow, S. A. & Higuchi, H. *Nature* **406**, 913–916 (2000).
2. Rice, S. *et al.* *Nature* **402**, 778–784 (1999).
3. Case, R. B., Rice, S., Hart, C. L., Ly, B. & Vale, R. D. *Curr. Biol.* **10**, 157–160 (2000).
4. Endow, S. A. *Nature Cell Biol.* **1**, E163–E167 (1999).

5. Hirose, K., Lowe, J., Alonso, M., Cross, R. A. & Amos, L. A. *Mol. Biol. Cell* **10**, 2063–2074 (1999).
6. Alonso, M. C., van Damme, J., Vandekerckhove, J. & Cross, R. A. *EMBO J.* **17**, 945–951 (1998).
7. Sablin, E. P. *et al.* *Nature* **395**, 813–816 (1998).
8. Crevel, I. M., Lockhart, A. & Cross, R. A. *J. Mol. Biol.* **273**, 160–170 (1997).
9. Pechatnikova, E. & Taylor, E. W. *Biophys. J.* **77**, 1003–1016 (1999).
10. Foster, K. A. & Gilbert, S. P. *Biochemistry* **39**, 1784–1791 (2000).
11. Stewart, R. J., Semerjian, J. & Schmidt, C. F. *Eur. Biophys. J.* **27**, 353–360 (1998).
12. Kozielewski, E., De Bonis, S., Burmeister, W. P., Cohen-Addad, C. & Wade, R. H. *Structure Fold. Des.* **7**, 1407–1416 (1999).
13. Lockhart, A. & Cross, R. A. *EMBO J.* **13**, 751–757 (1994).
14. Veigel, C. *et al.* *Nature* **398**, 530–533 (1999).

Quantum Hall effect

Composite fermions pair up

Nick Bonesteel

In 1956 Leon Cooper¹ published a simple ‘back of the envelope’ calculation that provided a key insight into the nature of superconductivity. Cooper was interested in the quantum-mechanical description of electrons in a metal. Electrons are fermions — particles which obey the Pauli exclusion principle forbidding two identical particles from being in the same quantum state. Because of this, at low temperatures and in the absence of any instability, a gas of electrons will form something called a Fermi sea. Cooper knew that although electrons normally repel each other, in a metal they can feel a net attraction due to tiny lattice vibrations (phonons). He then showed that if such an attraction is present, two electrons added to a Fermi sea will always form a ‘molecular’ bound state called a Cooper pair. This result was an important step towards the development of the full Bardeen–Cooper–Schrieffer (BCS) theory of superconductivity. On page 863 of this issue, Scarola *et al.*² present a modern version of Cooper’s calculation in a new and surprising context, that of the quantum Hall effect.

The quantum Hall effect is the modern offspring of an observation first made by Edwin Hall in 1879. The Hall effect occurs when a conductor carrying an electrical current is put in a magnetic field and a voltage develops perpendicular to both the current and the field. The sign of the Hall voltage reveals whether negative or positive charges are carrying the current. Later experiments at very low temperatures and in high magnetic fields led to a quantum version of this effect in a two-dimensional conductor. Under these conditions, the gas of electrons is trapped in two dimensions and the voltage changes in quantized steps (rather than smoothly) as the magnetic field is increased.

Physicists often measure the strength of a magnetic field in natural units called magnetic flux quanta. A two-dimensional gas of

electrons in a magnetic field is then characterized by its ‘filling fraction’ ν — the number of electrons per flux quantum. The quantum Hall effect can occur when the filling fraction is an integer, $\nu = p$ (integer

quantum Hall effect), or a rational fraction, $\nu = p/q$ (fractional quantum Hall effect).

When the quantum Hall effect occurs, the electron gas becomes incompressible — the density of particles becomes fixed or ‘pinned’ at a specific value. For the integer effect, this pinning can be understood in terms of non-interacting electrons obeying the Pauli exclusion principle. But for the fractional effect, the Coulomb repulsion between negatively charged electrons must be included in the calculation. As shown by Robert Laughlin, this repulsion can lead to subtle correlations in the electrons’ behaviour that allow them to condense into an incompressible quantum fluid at certain rational filling fractions.

New insight into the nature of the quantum Hall effect was provided when Jain introduced the composite fermion³. This can be thought of as an electron bound to an even number of magnetic flux quanta (or vortices). Like electrons, composite fermions obey the Pauli exclusion principle. But unlike electrons, composite fermions see an effective magnetic field that is much smaller than the applied magnetic field. Accordingly,

Behavioural ecology

Eat me!

The parasitic fluke *Microphallus piriformes* has a problem. To complete its life cycle it needs to travel between two hosts: the rough periwinkle (*Littorina saxatilis*, pictured, a seashore mollusc) and the herring gull. In the normal run of things, these species have little to do with each other. But the cunning parasite has a way of making introductions.

By comparing the behaviour of infested and healthy periwinkles, Helen McCarthy and colleagues have discovered that *M. piriformes* seems to bend the periwinkles’ behaviour to its own ends. In both the laboratory and field experiments on Muck Island, Scotland, parasitized periwinkles showed a greater tendency to crawl upwards — into positions where they are more visible to gulls, and presumably more likely to be eaten by them — than their healthier counterparts (*Anim. Behav.* **59**, 1161–1166; 2000).

To reach their elevated positions, the suicidal periwinkles reduce their



amounts of horizontal and downwards travel. It is the direction, rather than the amount, of movement that changes. Infected animals also alter their responses to the tide, moving upwards as it rises — healthy periwinkles do the opposite.

This change in behaviour happens only when the infection is mature and the fluke is ready to switch hosts. In the early stages, infected periwinkles behave normally; after all, the parasite doesn’t want its home to perish from desiccation or predation too early. The parasite’s timing also seems designed to bring periwinkles and gulls into contact during the summer, when gulls are gathered at their breeding colonies. The

proportion of parasitized periwinkles is much greater near gull breeding colonies than at their foraging sites, although McCarthy *et al.* have yet to show whether infested periwinkles are actually more likely to find their way into a gull’s stomach.

Nor are vertebrates immune to this form of parasitic trickery. A paper by Manuel Berdoy and colleagues, published earlier this month (*Proc. R. Soc. Lond. B* **267**, 1591–1594; 2000), shows that rats infected with the protozoan *Toxoplasma gondii*, which they catch from eating cat faeces, become more curious and less fearful. This makes them easier prey for cats — enabling *Toxoplasma* to get back to its preferred host. **John Whitfield**

HEATHER ANGEL

Jain showed that the fractional quantum Hall effect for electrons can be seen as an integer quantum Hall effect for composite fermions.

A striking consequence of this theory is that for a filling fraction $\nu = 1/2$, the effective magnetic field seen by composite fermions is zero⁴. It is then possible for composite fermions to form a Fermi sea, analogous to the Fermi sea of electrons in an ordinary metal. This description is consistent with the observation that the $\nu = 1/2$ state is compressible — that is, it does not show the quantum Hall effect. Other experiments around $\nu = 1/2$ have revealed behaviour characteristic of a two-dimensional metal in zero field⁵.

Composite fermion theory predicts that at a filling fraction of $\nu = 5/2$, the effective magnetic field should also be zero. But experiments show a quantum Hall effect at $\nu = 5/2$ (refs 6, 7), suggesting that this is not a simple metallic state. An intriguing possibility^{8,9}, and one that has recently gathered theoretical support^{10,11}, is that at $\nu = 5/2$ a Fermi sea of composite fermions forms but is susceptible to Cooper pairing and so becomes a composite fermion 'superconductor'.

Why would such pairing lead to a quantum Hall effect? One way to understand this is in terms of the Meissner effect — the fact that superconductors expel magnetic fields. If a Fermi sea of composite fermions does form a 'superconducting' state of Cooper pairs, then, because adding more electrons to the system is equivalent to applying a magnetic field, any added particles will be expelled. This leads to just the sort of pinning density needed for the quantum Hall effect.

With this in mind, Scarola *et al.*² have carried out a Cooper calculation for composite fermions. Unlike Cooper's original work, this was too difficult to do on the back of an envelope. Instead, Scarola *et al.* explicitly constructed quantum wavefunctions for the relevant composite fermion states and calculated their energies numerically. Using this technique they could start with a Fermi sea of composite fermions, introduce (or remove) two composite fermions, and compute the binding energy of the added (or removed) pair. Their results show that at $\nu = 1/2$ the composite fermions do not bind, but at $\nu = 5/2$ they do, suggesting Cooper pairing.

Given that the only interaction between the particles in this model is their Coulomb repulsion, it is natural to wonder about the origin of the attractive interaction causing Cooper pairing. Scarola *et al.* attribute it to an 'overscreening' of the Coulomb repulsion that occurs when the negatively charged electrons bind to the positively charged magnetic vortices to become composite fermions. Roughly speaking, because the short-range part of the effective Coulomb repulsion between electrons at $\nu = 5/2$ is weaker than at $\nu = 1/2$, Scarola *et al.* argue that it is possible for there to be a residual

attractive interaction between composite fermions at $\nu = 5/2$ but not at $\nu = 1/2$.

Scarola *et al.*'s calculation is a beautiful demonstration that a Fermi sea of composite fermions can form Cooper pairs, and gives some insight into why such pairing occurs at $\nu = 5/2$ but not $\nu = 1/2$. Like Cooper's 1956 calculation, this is an important step forward in our understanding of the $\nu = 5/2$ state, but it is not yet a full BCS theory of composite fermion pairing. It is possible that such a theory, when it arrives, may help explain another phenomenon in which pairing arises from purely repulsive interactions — high-temperature superconductivity. ■

Nick Bonesteel is in the Department of Physics and National High Magnetic Field Lab, Florida State University, Tallahassee, Florida 32310, USA.

e-mail: bonesteel@magnet.fsu.edu

1. Cooper, L. N. *Phys. Rev.* **104**, 1189–1190 (1956).
2. Scarola, V. W., Park, K. & Jain, J. K. *Nature* **406**, 863–865 (2000).
3. Jain, J. K. *Phys. Rev. Lett.* **63**, 199–202 (1989).
4. Halperin, B. I., Lee, P. A. & Read, N. *Phys. Rev. B* **47**, 7312–7343 (1993).
5. Willett, R. L. *Adv. Phys.* **46**, 447–544 (1997).
6. Willett, R. L. *et al. Phys. Rev. Lett.* **59**, 3846–3849 (1987).
7. Pan, W. *et al. Phys. Rev. Lett.* **83**, 3530–3533 (1999).
8. Moore, G. & Read, N. *Nucl. Phys. B* **360**, 362–396 (1991).
9. Greiter, M., Wen, X. G. & Wilczek, F. *Nucl. Phys. B* **374**, 567–614 (1992).
10. Morf, R. *Phys. Rev. Lett.* **80**, 1505–1508 (1998).
11. Rezayi, E. H. & Haldane, F. D. M. *Phys. Rev. Lett.* **84**, 4685–4688 (2000).

Neurobiology

Frazzled precision guides axons

Roman J. Giger and Alex L. Kolodkin

A functional nervous system depends on an intricate, finely tuned network of neuronal connections. During development, axons and dendrites, the specialized projections of a neuron, extend from the neuronal cell body over what are often long distances to establish contact with their targets. To ensure accurate targeting, axons and dendrites use attractive and repulsive guidance cues along their trajectories. Using the fruitfly *Drosophila melanogaster* as a model system, Hiramoto and colleagues, writing on page 886 of this issue¹, offer mechanistic insight into how secreted, diffusible cues can

provide precise guidance information. The authors show that a member of the Netrin family of diffusible guidance proteins is captured far from its site of synthesis and presented to approaching axons by its binding partner, a protein called Frazzled. The Netrin protein thereby instructs growing axons to follow a precise trajectory within the central nervous system (CNS).

The neuronal growth cone — a subcellular structure at the leading end of extending nerve fibres — integrates a myriad of guidance cues, and responds by dynamically adjusting the overall direction in which

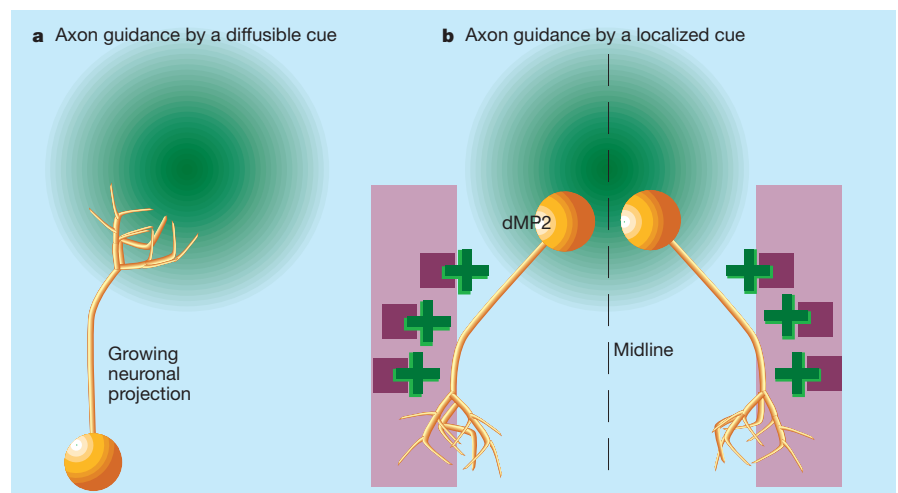


Figure 1 Guiding axon trajectories in developing embryos. a, Some axons respond to smooth gradients of attractive guidance cues (green) *in vitro* and *in vivo*. Presenting guidance cues in this way does not allow for abrupt changes in axon trajectory. b, If a secreted attractive guidance cue is selectively localized to an intermediate axonal target, guidance events far from the site at which the cue is made can be controlled accurately. For example, in the *Drosophila* embryo, dMP2 neurons extend laterally away from the midline and then turn posteriorly in a dorsolateral region of the central nervous system (CNS). This turn requires the chemoattractant Netrin (green plus symbols). The Netrin receptor Frazzled (dark purple) is localized to the surface of axons (light purple) in the dorsolateral region of the CNS. Hiramoto *et al.*¹ show that Frazzled sequesters secreted Netrin onto these lateral neurons, allowing the trajectory of the dMP2 axon to be precisely redirected.

projections extend². Some guidance cues are produced near the advancing growth cone, are tethered to the surface of nearby cells or to the extracellular environment, and can locally steer growth cones. Other, secreted and diffusible, guidance cues are thought to act over long distances by forming gradients that can either attract (Fig. 1a) or repel extending neuronal projections. But little is known about how these cues are presented to neuronal projections, or how their range of action is regulated during neural development.

Netrins are potent secreted guidance cues^{3,4}. In organisms from worms to mammals, Netrins attract axons from specific neuronal populations that cross from one side of the body to the other, towards the midline of the CNS. Long-range attraction by Netrin is mediated by receptors of the Deleted in Colorectal Cancer (DCC) protein family, which includes the *Drosophila* protein Frazzled. Intriguingly, Netrins can repel other populations of neurons that are directed away from the midline. Long-range repulsion by Netrin is mediated by a receptor complex that includes DCC and a member of the UNC-5 family of cell-surface-located receptors^{5,6}. Netrin receptors can be found on the surface of the growing axon and, on binding secreted Netrins, they send signals inside the axon to tell it which way to go. But in *Drosophila*, Frazzled is sometimes required not on the growing axons, but rather in their intermediate targets. How might Frazzled be used here?

To tackle this problem, Hiramoto *et al.*¹ looked at the distribution of Netrin in the developing fly nervous system. There are two Netrins in *Drosophila*, Netrin-A and Netrin-B. Both are strongly expressed in the CNS midline, where they work to attract axons^{7,8}. Interestingly, Hiramoto *et al.* find that a substantial fraction of Netrin protein also localizes to axons in a dorsolateral CNS region, far from where Netrin is made. This dorsolateral region also contains abundant Frazzled protein; indeed, Frazzled and Netrin are found on the same dorsolateral axons. This localization of Frazzled requires its cytoplasmic domain. A closer look revealed dramatic changes in Netrin distribution either in the absence of Frazzled or following the forced expression of Frazzled in the wrong places. So Frazzled plays a key part in capturing and localizing Netrin protein at specific sites in the fly CNS.

Does this precise Netrin localization serve to guide axons? Hiramoto *et al.* looked at CNS neurons called dMP2 neurons, which are known to require Netrin and Frazzled for guidance, and determined how much these neurons rely on dorsolaterally localized Netrin. In wild-type *Drosophila* embryos, axons from dMP2 neurons first grow laterally away from the midline, then turn posteriorly to grow parallel to the body

axis within the CNS (Fig. 1b). Hiramoto *et al.* find that, in embryos lacking Netrin or Frazzled, this characteristic turning behaviour of dMP2 axons is altered: they extend too far laterally and do not turn posteriorly. In embryos lacking Frazzled, this altered behaviour can be corrected by selectively expressing Frazzled in lateral neurons, but not in dMP2 neurons themselves, consistent with the observation that dMP2 neurons do not express Frazzled. Experiments using altered Frazzled proteins show that the ability of Frazzled to guide dMP2 axons and localize Netrin does not require Frazzled's cytoplasmic domain.

So, rather than serving only as a Netrin-binding signalling protein on the surface of growing neurons, Frazzled, and possibly other DCC-family members, also helps Netrin to steer axons that do not themselves express Frazzled. The extracellular domain of Frazzled captures and immobilizes Netrin on regions encountered by extending axons (Fig. 1b). This process does not require intracellular signalling events triggered by the binding of Netrin to Frazzled. It is well established that axons can respond both *in vitro* and *in vivo* to smooth gradients of guidance cues. But Hiramoto *et al.*¹ have uncovered an elegant mechanism that avoids problems inherent in using the free diffusion of cues to precisely control guidance events at a distance.

These results imply that an as yet uncharacterized receptor on the growing axon must be responding to Netrin. Netrin appears to be acting as an attractant, so *Drosophila* UNC-5 seems an unlikely candidate for this receptor. And does Netrin alone, or a Netrin–Frazzled complex, function as the ligand for the guidance events described by Hiramoto and colleagues? Is the Frazzled-bound form of Netrin active, or does Netrin signal only when released from Frazzled, subject to the off-rate of this reaction? The possibility that a Netrin–Frazzled complex may be transported to the final site of Netrin accumulation also deserves serious consideration. Finally, given that other secreted guidance cues unrelated to Netrin may be similarly localized, we must be prepared to accept new roles for old guidance receptors. ■

Roman J. Giger and Alex L. Kolodkin are in the Department of Neuroscience, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.
e-mail: kolodkin@jhmi.edu

1. Hiramoto, M., Hirami, Y., Giniger, E. & Hotta, Y. *Nature* **406**, 886–889 (2000).
2. Mueller, B. K. *Annu. Rev. Neurosci.* **22**, 351–388 (1999).
3. Chisholm, A. & Tessier-Lavigne, M. *Curr. Opin. Neurobiol.* **9**, 603–615 (1999).
4. Culotti, J. G. & Merz, D. C. *Curr. Opin. Cell Biol.* **10**, 609–613 (1998).
5. Hamelin, M., Zhou, M., Su, M. W., Scott, I. M. & Culotti, J. G. *Nature* **364**, 327–330 (1993).
6. Hong, K. *et al. Cell* **97**, 927–941 (1999).
7. Harris, R., Sabatelli, L. M. & Seeger, M. A. *Neuron* **17**, 217–228 (1996).
8. Mitchell, K. J. *et al. Neuron* **17**, 203–215 (1996).

Daedalus

Tension in miniature

The smaller a component, the stronger it tends to be — which explains the value of fine-fibre reinforcements. A very fine fibre can carry so much tension as to reach an almost explosive strain-energy density. Daedalus is now devising a fine powder whose particles are tense with locked-in stress.

Stress, he remarks, can accumulate during crystallization. A needle crystal, growing in solution and bent while it is still very narrow, will develop intense stress as it thickens. Successive sleeves of new molecules are deposited under compression on its inner radius and under dilation on its outer one. If it cannot straighten, the accumulating stress will snap it. So, says Daedalus, grow a needle-crystal compound on a micro-scale circular seed, and it will thicken into a tiny ring, tense with stress. A small shock will crack it; it will snap straight, releasing its energy ballistically as flying fragments.

DREADCO chemists are now trying it. Their circular seeds are plasmids, ring-polymer molecules, and ring-shaped structures made by coating a micro-fine wire, stretching it to shatter the coating into tiny rings, and dissolving it to release them. The chemistry is still unfocused and exploratory, but some promising systems should emerge soon. The final product will look like any ordinary white powder: flour, salt or cement. But it will be tense with strain energy.

DREADCO's Stressed Powder, crystallized almost to spontaneous explosion, will be highly dangerous. Its fiercely screwed-up particles will need a thin inert coating to prevent the mutual abrasion of handling or pouring from setting them off. It will form a novel physical explosive. Fired by a sudden shock, it will release no chemical fumes, just a violent hot blast of abrasive fragments. A weaker grade of Stressed Powder, removed from the crystallizing bath earlier, will have wider uses. It will make a splendid active abrasive, cutting through the hardest workpiece by forceful local energy release. It will transform the technology of surface abrasion, from etching to paint-removal to dental hygiene. Grit-blasted at a workpiece from an air-nozzle or sunk into it from a flexible pipe, Stressed Powder will penetrate like a drill. It will solve the age-old problem of making a deep hole, maybe curved or of non-circular cross-section, in a hard solid. Electricians facing the once intractable task of putting new wiring into an old stone building will bless the name of DREADCO.

David Jones

Obituary

Elsie Widdowson (1906–2000)

For seven years before her death on 14 June 2000, Elsie Widdowson was the most highly honoured woman scientist in Britain. She is best known now for her work in nutrition, but, as she pointed out when I was preparing her biography, published in 1993: "Nutrition as a subject did not exist when I started. I have been a chemist, biochemist, plant physiologist, medical researcher and a physiologist." When pushed, she conceded that the significance of genetic influences was the most important unanswered question in nutrition. But — sorry, *Nature* readers — she had a rather sceptical view of the enormous research effort into molecular biology, saying: "Why are they so keen to learn more and more about less and less?"

She gained her first degree in chemistry at Imperial College, London, in 1928. The turning point in her life came when, at the end of her PhD, in which she analysed developing apples for different carbohydrates, she decided to go to King's College Hospital in London to learn about large-scale cooking before starting a dietetics course. "I didn't really want to be a dietitian but jobs in research were hard to come by for beginners in the early 1930s," she said. A trip to the kitchens brought her into contact with Robert McCance, who was analysing the carbohydrate content of plant foods to plan diets for diabetics. In an inauspicious start, she had the cheek to tell him that he was using the wrong methods to determine the sugar content of apples!

But no matter, this first encounter led to a scientific partnership that was to last 60 years. McCance and Widdowson helped to shape wartime rationing and the British loaf, paved the way for later work on the damage that poor childhood nutrition does to adult health, and provided the core for almost every nutritional database in use in the world today. Their book *The Chemical Composition of Foods* became so universal on its publication in 1940 that their names passed into the title, linking them forever with the field they had made their own. The sixth edition of *McCance and Widdowson's The Composition of Foods* will be dedicated to Widdowson when it is published later this year.

Armed with these comprehensive 'food tables', Elsie realized that she was in a strong position to calculate the energy and nutrient intakes of individual adults. Her results, published in 1936, highlighted the large variation in these intakes between one individual and another of the same



Nutritionist who helped to shape wartime rationing

sex and age. Later she confirmed this finding by studying 1,000 children between the ages of 1 and 18 years. What she then identified as 'nutritional individuality' has now become the basis for detailed 'functional genomics' studies.

Widdowson shared McCance's enthusiasm for self-experimentation, which started with their investigations into the way that the body handles the economy (absorption and excretion) of minerals. They tried to persuade the body to excrete iron by injecting quite large quantities intravenously into themselves. She thought it was a pity to go to all the trouble of balanced experiments just for iron, so they had injections of several other minerals at the same time. All went well until "that dreadful Saturday afternoon (in 1938) when we had injected some strontium lactate into each other just before lunch to find out how we would excrete it. After about 40 minutes, we both began to have the most dreadful pyrogen reactions. We lay rolling about on that floor in misery." Fortunately for science, they were found.

When war was declared, McCance and Widdowson, who by then were working in Cambridge, decided to do research designed directly to help the war effort by establishing how far home-produced food could meet dietary needs. They drew up a scheme based on a weekly ration of just one egg, 1 lb meat or fish, 6 oz fruit, 5 oz sugar, 4 oz cheese and 4 oz fat. Brown bread and vegetables, including potatoes,

were not rationed, and a quarter of a pint of milk was allowed daily. But could anyone exist on that? To show that they could, McCance, Widdowson and several colleagues followed the diet to the letter for three months. They existed largely on bread, cabbage and potatoes, a combination that left them permanently full and strong. But would these experimental rations maintain fitness when the body was tested by much more severe conditions? They would find out.

Immediately after Christmas 1939 they set out for the Lake District. They hiked energetically across the fells for ten days with little to eat but bread. Their longest day in that fortnight was the one in which they covered 36 miles with 7,000 feet of climbing, at an average speed of 3.2 miles an hour including stops. It was a truly pioneering experiment, fastidiously recorded and confronting big questions in the practical, do-it-yourself way that was to typify their best work.

Throughout the war, more experiments on bread followed, the aim being to find out whether the more nutritious brown loaves being forced on a wartime population that preferred white interfered with calcium absorption. The results proved clear-cut: brown bread did reduce calcium absorption compared with white bread. As a result, every British loaf except wholemeal, which most needs it, is to this day fortified with calcium carbonate — wholemeal bread was excluded in deference to the pure-food enthusiasts who were vociferous even then.

McCance retired in 1968 but Widdowson continued to work for another 30 years, emerging at last from his shadow and, to the delight of her friends, achieving full public recognition in her own right. Innumerable honours and medals followed.

Elsie loved helping and encouraging young people. Her thatched cottage was a haven of tranquillity, replete with several pet cats, an extensive apple orchard and a vegetable garden, the latter providing practical nourishment for her many visitors.

What advice would Widdowson give to a young scientist? "If your results don't make physiological sense, think and think again! You may have made a mistake (in which case own up to it) or you may have made a discovery. Above all, treasure your exceptions. You will learn more from them than all the rest of your data."

Margaret Ashwell

Margaret Ashwell is at Ashwell Associates, Ashwell Street, Ashwell, Hertfordshire SG7 5PZ, UK.
e-mail: ashwell@compuserve.com

Navigation in a small world

It is easier to find short chains between points in some networks than others.

The small-world phenomenon — the principle that most of us are linked by short chains of acquaintances — was first investigated as a question in sociology^{1,2} and is a feature of a range of networks arising in nature and technology^{3–5}. Experimental study of the phenomenon¹ revealed that it has two fundamental components: first, such short chains are ubiquitous, and second, individuals operating with purely local information are very adept at finding these chains. The first issue has been analysed^{2–4}, and here I investigate the second by modelling how individuals can find short chains in a large social network.

I have found that the cues needed for discovering short chains emerge in a very simple network model. This model is based on early experiments¹, in which source individuals in Nebraska attempted to transmit a letter to a target in Massachusetts, with the letter being forwarded at each step to someone the holder knew on a first-name basis. The networks underlying the model follow the ‘small-world’ paradigm³: they are rich in structured short-range connections and have a few random long-range connections.

Long-range connections are added to a two-dimensional lattice controlled by a clustering exponent, α , that determines the probability of a connection between two nodes as a function of their lattice distance (Fig. 1a). Decentralized algorithms are studied for transmitting a message: at each step, the holder of the message must pass it across one of its short- or long-range connections; crucially, this current holder does not know the long-range connections of nodes that have not touched the message. The primary figure of merit for such an algorithm is its expected delivery time T , which represents the expected number of steps needed to forward a message between a random source and target in a network generated according to the model. It is crucial to constrain the algorithm to use only local information — with global knowledge of all connections in the network, the short-chain can be found very simply⁶.

A characteristic feature of small-world networks is that their diameter is exponentially smaller than their size, being bounded by a polynomial in $\log N$, where N is the number of nodes. In other words, there is always a very short path between any two nodes. This does not imply, however, that a decentralized algorithm will be able to discover such short paths. My central finding is that there is in fact a unique value of the exponent α at which this is possible.

When $\alpha = 2$, so that long-range connections

follow an inverse-square distribution, there is a decentralized algorithm that achieves a very rapid delivery time; T is bounded by a function proportional to $(\log N)^2$. The algorithm achieving this bound is a ‘greedy’ heuristic: each message holder forwards the message across a con-

nection that brings it as close as possible to the target in lattice distance. Moreover, $\alpha = 2$ is the only exponent at which any decentralized algorithm can achieve a delivery time bounded by any polynomial in $\log N$; for every other exponent, an asymptotically much larger delivery time is required, regardless of the algorithm employed (Fig. 1b).

These results indicate that efficient navigability is a fundamental property of only some small-world structures. The results also generalize to d -dimensional lattices for any value of $d \geq 1$, with the critical value of the clustering exponent becoming $\alpha = d$. Simulations of the greedy algorithm yield results that are qualitatively consistent with the asymptotic analytical bounds (Fig. 1c).

In the areas of communication networks⁷ and neuroanatomy⁸, the issue of routing without a global network organization has been considered; also in social psychology and information foraging some of the cues that individuals use to construct paths through a social network or hyper-linked environment have been discovered^{9,10}. Although I have focused on a very clean model, I believe that a more general conclusion can be drawn for small-world networks — namely that the correlation between local structure and long-range connections provides critical cues for finding paths through the network.

When this correlation is near a critical threshold, the structure of the long-range connections forms a type of gradient that allows individuals to guide a message efficiently towards a target. As the correlation drops below this critical value and the social network becomes more homogeneous, these cues begin to disappear; in the limit, when long-range connections are generated uniformly at random, the result is a world in which short chains exist but individuals, faced with a disorienting array of social

contacts, are unable to find them.

Jon M. Kleinberg
Department of Computer Science, Cornell University, Ithaca, New York 14853, USA

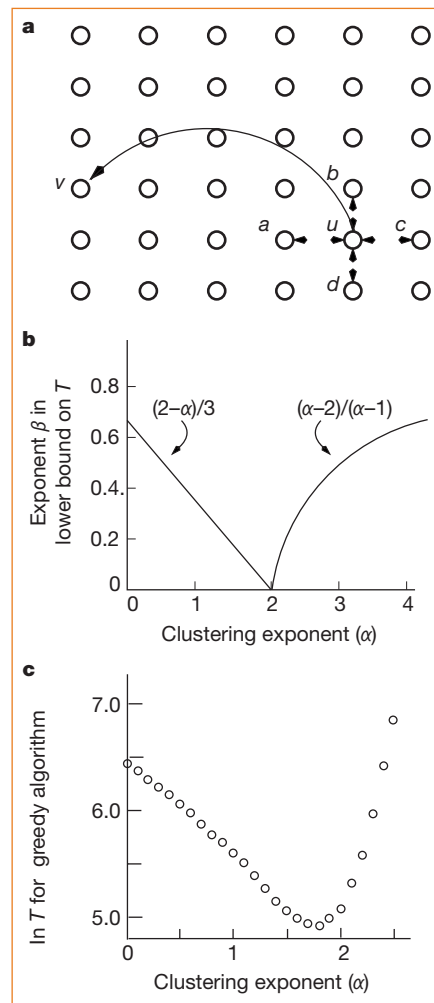


Figure 1 The navigability of small-world networks. **a**, The network model is derived from an $n \times n$ lattice. Each node, u , has a short-range connection to its nearest neighbours (a , b , c and d) and a long-range connection to a randomly chosen node, where node v is selected with probability proportional to $r^{-\alpha}$, where r is the lattice (‘Manhattan’) distance between u and v , and $\alpha \geq 0$ is a fixed clustering exponent. More generally, for $p, q \geq 1$, each node u has a short-range connection to all nodes within p lattice steps, and q long-range connections generated independently from a distribution with clustering exponent α . **b**, Lower bound from my characterization theorem: when $\alpha \neq 2$, the expected delivery time T of any decentralized algorithm satisfies $T \geq cn^\beta$, where $\beta = (2 - \alpha)/3$ for $0 \leq \alpha < 2$ and $\beta = (\alpha - 2)/(\alpha - 1)$ for $\alpha > 2$, and where c depends on α , p and q , but not n . **c**, Simulation of the greedy algorithm on a $20,000 \times 20,000$ toroidal lattice, with random long-range connections as in **a**. Each data point is the average of 1,000 runs.

1. Milgram, S. *Psychol. Today* **1**, 61–67 (1967).
2. Kochen, M. (ed.) *The Small World* (Ablex, Norwood, NJ, 1989).
3. Watts, D. & Strogatz, S. *Nature* **393**, 440–442 (1998).
4. Albert, R. et al. *Nature* **401**, 130–131 (1999).
5. Adamic, L. in *Proc. 3rd European Conference on Digital Libraries* (eds Abiteboul, S. & Vercoustre, A.-M.) 443–452 (Springer Lecture Notes in Computer Science, Vol. 1696, Berlin, 1999).
6. Cormen, T., Leiserson, C. & Rivest, R. *Introduction to Algorithms* (McGraw-Hill, Boston, 1990).
7. Peleg, D. & Upfal, E. *J. Assoc. Comput. Machinery* **36**, 510–530 (1989).
8. Braitenberg, V. & Schüz, A. *Anatomy of the Cortex* (Springer, Berlin, 1991).
9. Killworth, P. & Bernard, H. *Social Networks* **1**, 159–192 (1978).
10. Piroli, P. & Card, S. *Psychol. Rev.* **106**, 643–675 (1999).

Ecology

Nonlinearity and the Moran effect

The study of synchronization phenomena in ecology is important because it helps to explain interactions between population dynamics and extrinsic environmental variation^{1–5}. Grenfell *et al.*¹ have examined synchronized fluctuations in the sizes of two populations of feral sheep which, although situated on close but isolated islands, were nevertheless strongly correlated (observed value of the population correlation, r_p , 0.685). Using a nonlinear threshold model, they argue that this level of population correlation could only be explained if environmental stochasticity was correlated between the islands, with the environmental correlation, r_e , higher than 0.9 “on average” (Fig. 1a). This unusually high environmental correlation is far greater than would be predicted by the Moran effect², which states that the population correlation will equal the environmental correlation in a linear system. Grenfell *et al.*¹ imply that a simple nonlinearity in population growth can mask or even destroy the Moran effect^{1,3,4}. Here we show that these surprising results are an artefact of the techniques used to measure noise correlations and synchronization.

In their simulations¹, Grenfell *et al.* based all correlation calculations on a time series of $n=800$ model samples, which is far larger than the $n=18$ pairs of observed sheep-population data. To obtain a correct comparison of the model with their observed results, we repeated their calculations, but examined only $n=18$ pairs of simulated data samples for each model run; these smaller sample sizes should tend to increase statistical variability in the analysis.

Our study of an ensemble of simulation runs revealed that only a very weak environmental correlation r_e (calculated as for Fig. 1b) is required for the model populations frequently to achieve a correlation r_p greater than or equal to the observed correlation $r=0.685$. As Fig. 1d reveals, the observed correlation (or higher) can be found by chance alone, appearing in about one in 20 simulations for $r_e=0.3$, and then much more frequently as r_e increases.

The main conclusion of Grenfell *et al.*, that very strong environmental correlation ($r_e>0.9$) is responsible for the observed synchronized sheep-population fluctuations, therefore needs further consideration if a far smaller environmental correlation ($r_e=0.3$) can equally well explain the same dynamics. By failing to take into account the lengths of the ecological time series, their analysis may have misinterpreted the causes of synchrony in these feral sheep populations.

The Moran effect predicts that a plot of population correlation versus environmental correlation will yield a straight line of unit slope if the two synchronized systems are linear. By measuring r_e in the manner intended by Moran² (Fig. 1b, legend), we find that the nonlinear threshold model¹ provides results that scale very closely to the linear relationship (Fig. 1d), so the Moran effect survives. When r_p and r_e are plotted over the entire range of parameters used in ref. 1 (Fig. 1a and b, respectively), the two graphs are almost identical and $r_e \approx r_p$, a clear demonstration of the Moran effect.

Why then did Grenfell *et al.* fail to reproduce the Moran effect? In their approach, the environmental correlation r_e , as set by their model's control parameters, is often unrepresentative, as shown by the following argument. The environmental shocks actually experienced by the sheep populations

have, by construction, correlation r only when the populations are both above or below a fixed threshold level; but when one population is above threshold and the other below, a situation that can sometimes constitute half of any simulation run, the respective environmental shocks at the two islands are, again by construction, uncorrelated. These long periods of decorrelation can cause the true environmental correlation, r_e , calculated over the entire simulation run, to be markedly less than the value r defined by the model parameters. It is due to this misestimation of correlation (and not because of the presence of a nonlinear threshold^{1,3,4}) that Grenfell *et al.*¹ require an unusually high environmental correlation to yield the observed sheep synchrony.

Grenfell *et al.* have demonstrated the potential of using nonlinear statistical models for ecological investigations.

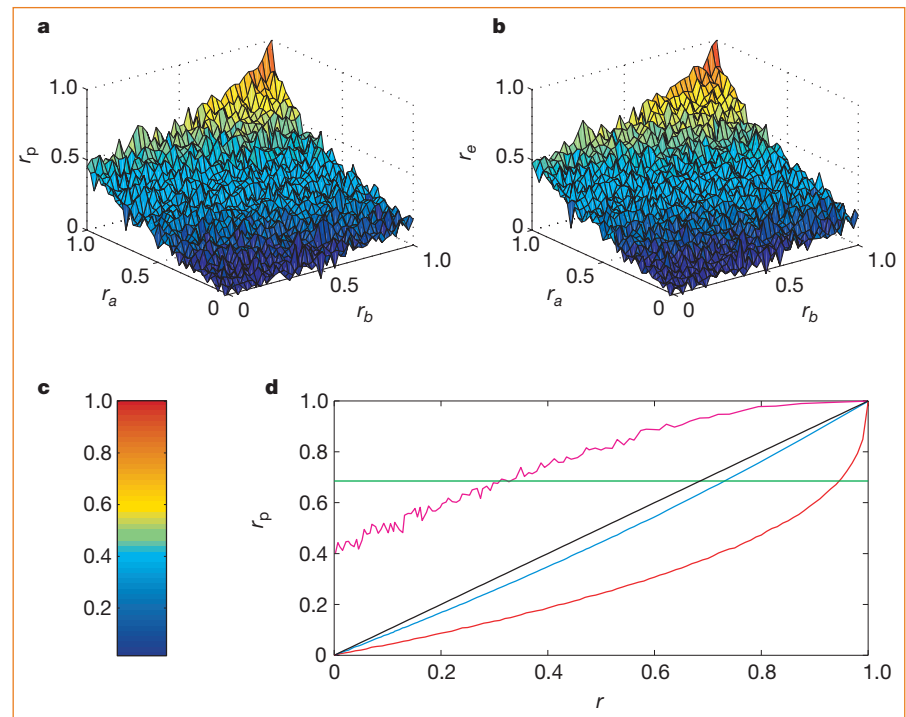


Figure 1 Simulation results using the nonlinear threshold model with density-dependent switching between two independent environmental noise components, as described by Grenfell *et al.*¹. **a**, Correlation between island sheep populations, r_p , as a function of the model's noise parameters r_a and r_b (Fig. 3a of ref. 1). The model¹ is set up so that r_a and r_b represent the correlation of the environmental noise affecting the two sheep populations when they are both either above or below threshold, respectively (see **c** for colour calibration). According to ref. 1, r_p equals the observed value of $r_o=0.658$ only if the environmental correlations r_a and r_b are larger than “0.9 on average”. **b**, True environmental correlation, r_e , as a function of the model parameters r_a and r_b . To calculate r_e , we stored all environmental shocks received by each model sheep population and then simply determined their correlation: hence r_e is viewed as the correlation coefficient of shocks directly experienced by the two model sheep populations. The surface for r_e is almost identical to that of r_p in **a**, demonstrating a Moran effect. **c**, Colour-coding of correlations depicted in **a** and **b**. **d**, Analysis of systems in which model environmental correlation parameters $r=r_a=r_b$ (that is, systems along the diagonal of parameter space in **a** and **b**). Red curve, cut through surface in **a** above the line $r=r_a=r_b$. The correlation between sheep populations r_p is plotted against the model's environmental correlation parameter r_e ; the deviation from the reference diagonal (black) seems to indicate the absence of a Moran effect, as argued in ref. 1. Blue curve, r_p vs r_e , where r_e (horizontal axis) is the true environmental correlation (see **b**). Note that this line is very close to the reference diagonal (black), where $r_p=r_e$, as predicted by the Moran effect. Magenta line, effects of small sample size ($n=18$) due to statistical scattering in an ensemble of 1,000 simulation runs: for each value of environmental correlation r_e (horizontal axis), 5% of the simulations have r_p values larger than the value plotted. Green horizontal line, the measured correlation between the two observed sheep populations is $r_o=0.685$. The point of intersection of the green and magenta lines yields the minimum value of the ‘true environmental correlation’, r_e , for which we can expect to find the observed sheep correlation by chance alone (that is, in at least 1 in 20 runs, or 5%, when $r_e=0.3$). Further details are available from the authors.

Refining these techniques should increase our understanding of the relation between environmental fluctuations and population dynamics, in the spirit of Moran.

Bernd Blasius, Lewi Stone

The Porter Super-Center for Ecological and Environmental Studies,
Department of Zoology, Tel Aviv University,
Ramat Aviv, Tel Aviv 69978, Israel
e-mail: lewi@lanina.tau.ac.il

1. Grenfell, B. T. *et al.* *Nature* **394**, 674–677 (1998).
2. Moran, P. A. P. *Aust. J. Zool.* **1**, 291–298 (1953).
3. Stenseth, N. C. & Chan, K.-S. *Nature* **394**, 620–621 (1998).
4. Hudson, P. J. & Cattadori, I. M. *Trends Ecol. Evol.* **14**, 1–2 (1999).
5. Blasius, B. *et al.* *Nature* **399**, 354–359 (1999).

Grenfell *et al.* reply — The Moran effect refers to systems of population dynamics that are linear: under these circumstances, the long-term correlation between population densities will be the same as the correlation between the random environmental perturbations. The Soay sheep exhibit significant nonlinearity in their density dependence (Fig. 2a of ref. 1). At low populations, numbers tend to increase exponentially, with mean growth rate $r = 0.24$, whereas at high densities (above a threshold of 1,172 animals), the population tends to decline, with mean $r = -0.29$. Thus, when populations on two adjacent islands are both above their thresholds, both will tend to decline, and when both are below their thresholds, both will tend to increase.

This immediately introduces a positive correlation between population dynamics on adjacent islands in the absence of any environmental forcing. When one population is above the threshold and one is below, the expectation is of no short-term correlation because the trends will tend to cancel out. Adding noise to the system has two effects. Correlated noise tends to push the dynamics into synchrony, because both populations tend to crash to low densities during the same years (for example, those with the most extreme winter weather). The nonlinearity means, however, that the same stochastic event could drive one population below the threshold but leave another population above it (for example, if initial population densities are sufficiently different). In this case, the two populations would experience different regimes of density dependence during the same year and synchrony would be reduced. Intuitively, then, in the presence of nonlinear density dependence, environmental forcing has to be stronger if it is to drive the populations into synchrony and keep them there.

Blasius and Stone have pointed out two problems with our analysis. First, they show that we had the two populations experiencing different — hence uncorrelated — noise during periods when the two populations were on opposite sides of the threshold. Correcting this mistake reduces the level

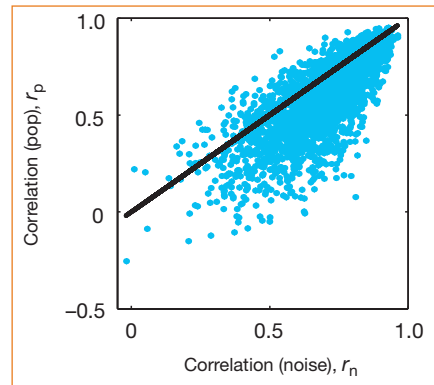


Figure 1 Scatter plot (blue) of inter-island population correlation (r_p) against true noise correlation (r_n) for 3,000 simulations (each of 18 time points) of the SETAR model, defined as in Table 1 of ref. 1, with the correction in noise realization proposed by Blasius and Stone. The black line indicates where population correlation equals the noise correlation (the expectation of the Moran effect). See ref. 2 for more details.

of noise correlation (r_n) required to produce the observed level of population correlation (r_p) ($r_p = 0.685$) from $r_n > 0.9$ to around 0.8 for large samples. This means that the extra-Moran effect is reduced, but not abolished.

Their second point is that, with realistically short time series (such as our 18 points), variability in the inter-population correlation coefficient generates a relatively high expectation of observing correlations higher than Moran, leading to inflated type-I errors. We have carried out further calculations with the corrected model that agree qualitatively with this. However, even short model simulations show the imprint of nonlinearity in their aggregate correlation structure — a strong downward bias in population correlation for a given level of noise correlation, compared to various linear null models (Fig. 1; for more details, see ref. 2). Thus, the impact of nonlinearity on population correlation is apparent in the collective behaviour of short simulations, as well as in individual realizations of the model's long-term dynamics.

There are several important directions for studies on the interactions between noise and determinism in population dynamics. Most important is an increase in the realism of the underlying model. The inclusion of age- and sex-structure effects is essential, because we know that animals of different ages and sexes experience markedly different patterns of mortality³. A further improvement would incorporate threshold density as a random variable rather than a constant (it is intraspecific competition for food that underlies the density dependence, and food supply determines the sheep density at which competition kicks in). This would allow for island-to-island differences in the response of food availability to environmental noise, so that islands with the same population densities could experience

different density-dependence regimes in the same year.

The ability to detect extra-Moran correlations depends critically on the balance between noise and density dependence (B. Blasius and L. Stone, personal communication), so any model refinement that explains more variability in terms of population processes will increase our powers of evaluation. Technically, developments in nonlinear time-series analysis need to encompass estimation of age⁴ and spatial heterogeneities, as well as the dissection of process noise from measurement error. This is particularly important for the relatively short time series found in ecology, where even linear time-series models can produce a complex range of correlation behaviours².

A thorough understanding of spatial dynamics can only come about once the interaction between correlated noise and nonlinear density dependence is understood through long-term ecological studies, combined with models.

B. T. Grenfell*, B. F. Finkenstädt†,

K. Wilson‡, T. N. Coulson§, M. J. Crawley||

*Zoology Department, University of Cambridge,
Downing Street, Cambridge CB2 3EJ, UK

†Department of Statistics, University of Warwick,
Coventry CV4 7AL, UK

‡Institute of Biological Sciences, University of
Stirling, Stirling FK9 4LA, UK

§Institute of Zoology, Zoological Society of London,
Regent's Park, London NW1 4RY, UK

||Imperial College, Silwood Park, Ascot SL5 7PY, UK

1. Grenfell, B. T. *et al.* *Nature* **394**, 674–677 (1998).
2. www.zoo.cam.ac.uk/zoostaff/matt/GROUP/Download/
3. Grenfell, B. T., Price, O. F., Albon, S. D. & Clutton-Brock, T. H. *Nature* **355**, 823–826 (1992).
4. Leirs, H. *et al.* *Nature* **389**, 176–180 (1997).

Physiology

An actively controlled heart valve

Vertebrate hearts typically have cardiac valves that are thin and leaf-like and which work passively, allowing blood to move forward during systole and preventing it from flowing back during diastole. Crocodilian hearts have nodules of connective tissue, resembling opposing knuckles, or cog-teeth^{1–3}, in the subpulmonary conus just proximal to the pulmonary valves. Here we show that these cog-teeth act in the estuarine crocodile *Crocodylus porosus* (Fig. 1) as a valve that regulates the flow of blood between the lungs and the systemic circulation in response to a β -adrenergic mechanism. To our knowledge, this is the first report of an actively controlled intra-cardiac valve in a vertebrate.

Among the vertebrates, only the mammals, birds and crocodilians have anatomi-

cally separate ventricles that allow the complete separation of systemic and pulmonary circulations. However, shunting blood away from the lungs and into the systemic circulation is possible in the crocodilian heart because the right ventricle gives rise not only to the pulmonary arteries but also to another major vessel, the left aorta (Fig. 2a). This allows deoxygenated right-ventricular blood to bypass the lungs and to be re-circulated into the systemic circulation (pulmonary-to-systemic shunt) — that is, blood is ejected into the left aorta instead of the pulmonary arteries. For a pulmonary-to-systemic shunt to develop, the pressure in the right ventricle must exceed the pressure in the left aorta^{4–8}.

The cog-teeth^{1–3} connective tissue nodules (Fig. 2a, b) are one of the intriguing anatomical features of the crocodilian heart. They project from the subpulmonary conus into the pulmonary outflow tract just proximal to the pulmonary leaf-like valves and fit snugly together during systole, reducing the diameter of the subpulmonary conus³. Indirect evidence^{9–12}, such as the spontaneous appearance of a biphasic pressure peak in the right ventricle, suggests that the cog-teeth might be an 'extra' valve mechanism that could help in initiating and regulating shunts.

We used a crocodile heart preparation⁶ perfused *in situ* to determine the function and possible regulation of this valve. The perfused heart generated a cardiac output similar to that recorded *in vivo* and worked at physiological pressures. Using an antagonist of the β -adrenergic receptor, sotalol, we initiated a shunt where about a third of the output from the right ventricle exited through the left aorta. Injecting adrenaline or isoprenaline into the right side of the heart reduced flow to the left aorta and eventually abolished the shunt, with flow being redirected into the pulmonary artery (Fig. 2c). This indicated that control of the shunt within the heart of *C. porosus* was

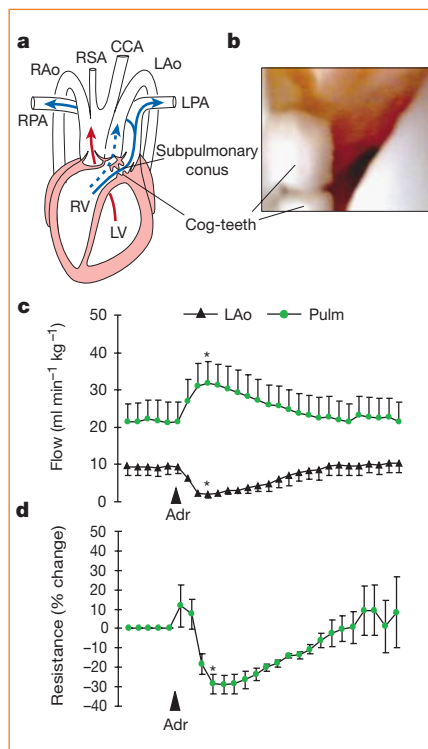


Figure 2 The 'cog-teeth' valve inside the heart of crocodilians and the flow and resistance from a perfused *Crocodylus porosus* heart preparation⁶. **a**, The crocodilian heart, outflow tract and major arteries, showing the position of the cog-teeth within the subpulmonary conus, proximal to the leaflet pulmonary valves and surrounded by cardiac muscle. During non-shunting conditions, blood is ejected from the left ventricle (LV) into the right aorta (RAo) (leading to the dorsal aorta), right subclavian artery (RSA) and the common carotid artery (CCA) (red arrow). The right ventricle (RV) ejects blood into the common pulmonary trunk which divides into the left and right pulmonary arteries (LPA and RPA) (blue arrow). During shunting, blood is ejected from the right ventricle into the left aorta (LAo) (dotted arrow). **b**, Angioscopic image³ of the 'cog-teeth' in the subpulmonary conus in the right ventricle; pointers show individual cogs. Photo courtesy of C. Löfman. **c**, The effect of a bolus injection of adrenaline (Adr, 0.3 ml 10⁻⁷ M) on pulmonary and left aortic flows, and **d**, on pulmonary outflow resistance, in a *C. porosus* perfused-heart preparation pretreated with the β -adrenoceptor-antagonist sotalol. Sotalol induced a shunt in the perfused heart preparation, despite the left-aortic output pressure being set at 1.5 kPa above the pulmonary artery output pressure. A bolus injection of adrenaline eliminated the shunt, with flow in the left aorta being diverted back into the pulmonary outflow tract. The bolus injection of adrenaline resulted in a significant reduction (28.9 $\pm$ 4.9%) in pulmonary outflow resistance (means $\pm$ s.e.m.; n = 6; asterisks indicate a significant difference at P < 0.05, Wilcoxon signed rank test).

mediated by an intrinsic β -adrenoceptor.

Neither heart rate nor the total power output generated by the heart preparation changed during the injection of adrenaline/isoprenaline, excluding the possibilities that the shunt was mediated by a change in the inotropic state of the heart or as the result of a change in stroke volume (Starling effect). The shunt induced by sotalol was the result of a large increase in resistance of the pulmonary outflow tract. The injected boluses of adrenaline and isoprenaline decreased the resistance of the pulmonary outflow tract, corresponding to the increased flow in the pulmonary artery (Fig. 2d). As the effects of the bolus injections wore off, the resistance of the pulmonary tract increased again and the shunt returned.

Our results demonstrate the presence and action of a regulatory intra-cardiac valve within the subpulmonary conus of the estuarine crocodile and indicate that the initiation of a shunt occurs when the β -adrenergic drive on the heart is low.

Craig E. Franklin*, **Michael Axelsson†**

*Department of Zoology and Entomology, University of Queensland, Brisbane, Queensland 4072, Australia

e-mail: cfranklin@zen.uq.edu.au

†Department of Zoophysiology, University of Göteborg, PO Box 463, Göteborg 405 30, Sweden

- van Mierop, L. H. S. & Kutsche, L. M. in *Cardiovascular Shunts: Phylogenetic, Ontogenetic and Clinical Aspects* (Munksgaard, Copenhagen, 1985).
- Webb, G. J. W. *J. Morphol.* **161**, 221–240 (1979).
- Axelsson, M., Franklin, C. E., Löfman, C. O., Nilsson, S. & Grigg, G. C. *J. Exp. Biol.* **199**, 359–365 (1996).
- Grigg, G. C. & Johansen, K. *J. Comp. Physiol.* **157**, 381–392 (1987).
- Shelton, G. & Jones, D. R. *J. Exp. Biol.* **158**, 539–564 (1991).
- Franklin, C. E. & Axelsson, M. *J. Exp. Biol.* **186**, 269–288 (1994).
- Axelsson, M., Nilsson, S. & Holm, S. *Am. J. Physiol.* **256**, R875–R879 (1989).
- Malvin, G. M., Hicks, J. W. & Green, R. E. *Am. J. Physiol.* **269**, R1133–R1139 (1995).
- White, F. *Copeia* **3**, 567–570 (1969).
- Greenfield, L. J. & Morrow, A. G. *J. Surg. Res.* **1**, 97–103 (1961).
- White, F. N. *Am. Zool.* **8**, 211–219 (1968).
- Jones, D. R. & Shelton, G. *J. Exp. Biol.* **176**, 247–269 (1993).

Erratum

A triclosan-resistant bacterial enzyme

Richard J. Heath, Charles O. Rock

Nature **406**, 145–146 (2000)

In Table 1, the third entry should have been *Streptococcus pneumoniae*, and not *Salmonella pneumoniae* as published.



Figure 1 The estuarine crocodile *Crocodylus porosus* has an unusual valve inside its heart that can shunt blood away from the lungs.

Transport, capture and exocytosis of single synaptic vesicles at active zones

D. Zenisek^{*†}, J. A. Steyer^{†‡} & W. Almers^{*}

^{*} Vollum Institute, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201, USA

[‡] Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

[†] These authors contributed equally to this work

To sustain high rates of transmitter release, synaptic terminals must rapidly re-supply vesicles to release sites and prime them for exocytosis. Here we describe imaging of single synaptic vesicles near the plasma membrane of live ribbon synaptic terminals. Vesicles were captured at small, discrete active zones near the terminal surface. An electric stimulus caused them to undergo rapid exocytosis, seen as the release of a fluorescent lipid from the vesicles into the plasma membrane. Next, vesicles held in reserve about 20 nm from the plasma membrane advanced to exocytic sites, and became release-ready 250 ms later. Apparently a specific structure holds vesicles at an active zone to bring v-SNAREs and t-SNAREs, the proteins that mediate vesicle fusion, within striking distance of each other, and then allows the triggered movement of such vesicles to the plasma membrane.

Neurons stimulated at a high frequency often suffer 'short-term depression' in that their release of transmitter diminishes from stimulus to stimulus^{1–5}, but recovers after several seconds of rest. In goldfish retinal bipolar neurons, triggering a sustained calcium influx elicits a fast phase of exocytosis followed by a slower phase^{6–9}. Both findings have been taken to indicate that a small pool of release-ready vesicles is rapidly exhausted and then replenished from larger reserve pools^{1–5}. Why are some vesicles release-ready and most not, and what determines how rapidly they become releasable? Functional studies generally provide no direct answers to these questions, as they assay only the last step in the secretory pathway, exocytosis, and must infer precursor reactions by kinetic modelling.

It would clearly be helpful to record signals from vesicles before exocytosis. This is possible when larger secretory vesicles are fluorescently labelled and imaged in live endocrine^{10–13} and epithelial cells^{14,15}. Also, synaptic vesicles may be stained selectively, most conveniently with the fluorescent lipid FM1-43 (ref. 16). So far, however, it has not been possible to image synaptic vesicles (the smallest membranous organelles found in any cell) *in vivo*, even when exocytosis led to quantized loss of fluorescence^{17,18}.

Here we have used goldfish retinal bipolar neurons, whose cell bodies are connected by short axons to giant terminals filled with $0.5\text{--}1 \times 10^6$ synaptic vesicles¹⁹. The terminals readily take up FM1-43 when a stimulus triggers exocytosis of synaptic vesicles, and release the dye during a second bout of stimulated exocytosis^{20,21}. As in other synapses^{16,17}, FM1-43 taken up as a consequence of exocytosis is thought to stain synaptic vesicles specifically. Presumably the dye is internalized in clathrin-coated vesicles, as they bud off the plasma membrane or its invaginations and then shed their coats to become synaptic vesicles^{18,22–24}.

Imaging single synaptic vesicles

Even when dye is applied during a brief stimulus to stain only about 1% of a terminal's synaptic vesicles, terminals under epifluorescence reveal little structure other than a uniform ring of fluorescence²¹. To image only the plasma membrane and the organelles directly beneath it, we illuminated stained terminals from the surface to a depth of less than 100 nm by total internal reflection (Fig. 1a). Punctate fluorescence appeared superimposed on a dim glow from the plasma membrane in places where a terminal adhered tightly to a glass coverslip (Fig. 1b). Fluorescent spots appeared only with dye applied while synaptic vesicles underwent exocytosis, and then only in the terminal and not in the adjacent axon or cell body (data not

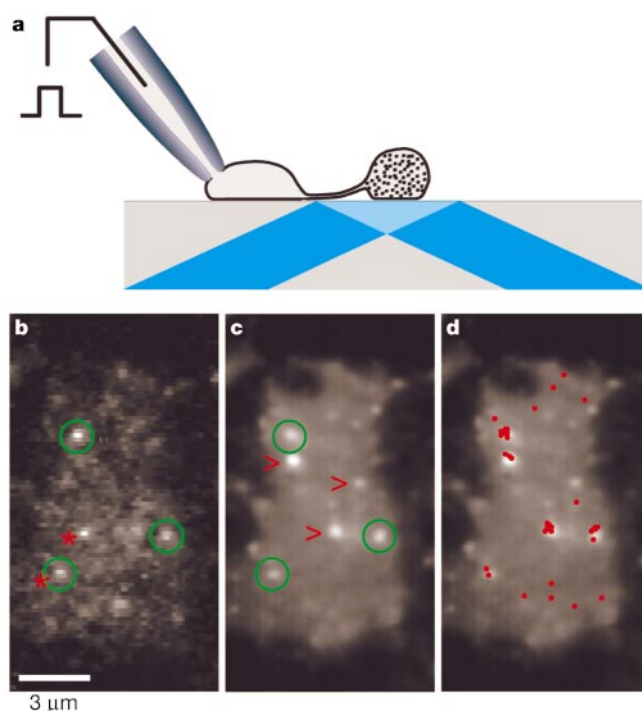


Figure 1 Active zones on the terminal of a bipolar neuron. **a**, Experimental set-up. FM1-43-loaded vesicles in the terminal were excited by the evanescent field generated when a laser beam underwent total internal reflection at the glass–cell interface. The soma of a bipolar neuron was voltage clamped. **b**, Single video frame showing a terminal with fluorescent spots superimposed on a mottled background. Among the brightest spots, two disappeared within the next two video frames (asterisks). Circles mark some of the active zones revealed in **c** and **d**. Background fluorescence represents FM1-43 remaining on the plasma membrane after washout. **c**, Average of 500 video frames (including **b**) collected during seven recording periods (1.33–2.5 s) during which electrical stimuli were applied. Bright spots mark sites occupied repeatedly and/or for extended periods by fluorescent vesicles; half of them were also occupied in **b** (circles), the rest were not (arrowheads). The background arose as in **b** and, to a lesser extent, from around a hundred vesicles colliding with the plasma membrane at random sites. **d**, Duplicate of **c** showing the location of exocytic events (red dots). Four bright, motionless spots, seen at the margins of each image for the entire experiment in **b** and **c** and probably representing FM1-43-stained debris, were removed for clarity.

shown). Apparently dye uptake is associated with the exocytosis of synaptic vesicles.

In video recordings (see movie 1 in Supplementary Information), some spots were visible only for a few video frames, as if caused by vesicles briefly entering the evanescent field as they hit and then bounced off the plasma membrane. Others were stationary, as if representing vesicles attached to the plasma membrane or to a nearby structure such as a synaptic ribbon. Stationary vesicles remained still for more than 0.5 s before they detached or were lost through exocytosis (see below). The frame in Fig. 1b shows vesicles engaged in both brief encounters (asterisks) and extended visits. To test whether vesicles occupy preferred sites, Fig. 1c superimposes 500 successive video images of the terminal. In this time-averaged view, bright spots represent sites where vesicles stayed for long periods near the plasma membrane or were repeatedly captured and lost. The near-uniform background glow is due mostly to small amounts of dye remaining in the plasma membrane and in small part to random collisions of vesicles with the plasma membrane. Of the six brightest sites, only three (circles) were occupied in Fig. 1b. Conversely, some spots in Fig. 1b appeared at sites not strongly stained in Fig. 1c, presumably because they were never or rarely occupied by other stained vesicles.

If the spots represent synaptic vesicles then they should release dye when Ca^{2+} enters the cell and triggers exocytosis. Square voltage steps were applied to the cell body and opened calcium channels in the terminal (average current 110 ± 10 pA, $n = 14$ cells). As predicted, spots brightened and spread as dye diffused laterally into the plasma membrane (Fig. 2a, see movies 2, 3 in Supplementary Information). The spread of dye indicates exocytosis, as it represents the release of material from vesicles. Figure 2c plots fluorescence against time for five fusing vesicles, both in a small circle around a spot and in a concentric annulus. Fluorescence increased first in the circle, then spread into the annulus (Fig. 2d) and finally declined as dye diffused away laterally into the plasma membrane. Fusion events of similar amplitude occurred stochasti-

cally during voltage steps (Fig. 2c, arrows), as if due to single synaptic vesicles. Most fusion events were concentrated in small active zones (Fig. 1d).

To determine the size of the fusing organelle, we measured the fluorescence spreading from vesicles into the plasma membrane (Fig. 3). Squares centred on a fusing vesicle were excised from video clips, and the fluorescence in each was plotted against its position on the cell surface. We used a frame acquired shortly before fusion (Fig. 3a) to calculate an image of the local background fluorescence (Fig. 3b, c); this background image (Fig. 3c) was subtracted from other frames in the video clip (Fig. 3d). Fluorescence brightened and spread after fusion. In contrast to the analysis shown in Fig. 2, it was possible in this analysis to sum all light from a vesicle's dye even while it spread over the plasma membrane. Summed light from background-subtracted images increased (Fig. 3e, circles); this was only in small part because dye released by exocytic events elsewhere in the cell caused a general brightening of the plasma membrane (triangles).

Allowance for this effect gave Fig. 3f. Light reached a plateau value ($7,656 \pm 413$ units, $n = 18$ vesicles) that we attribute to the vesicle's dye after its escape into the plasma membrane. Next, we stained plasma membrane to equilibrium at the dye concentration also used to stain vesicles; this was done in the absence of external Ca^{2+} to suppress vesicle turnover. Plasma membrane fluorescence was $3.07 \pm 0.30 \times 10^6$ units μm^{-2} ($n = 4$ cells). Comparison of the two values shows that the fusing organelles contributed the equivalent of $2,490 \pm 300$ nm² of stained plasma membrane. This area would cover a sphere with a diameter of 28.2 ± 1.6 nm, virtually identical to the diameter of synaptic vesicles in bipolar neurons (29 nm (ref. 19)). The result strongly supports the idea that we are observing single synaptic vesicles.

Vesicle capture and the kinetics of exocytosis

Figure 4a explores the kinetics of exocytosis by plotting the cumulative number of vesicle fusions against time during a voltage

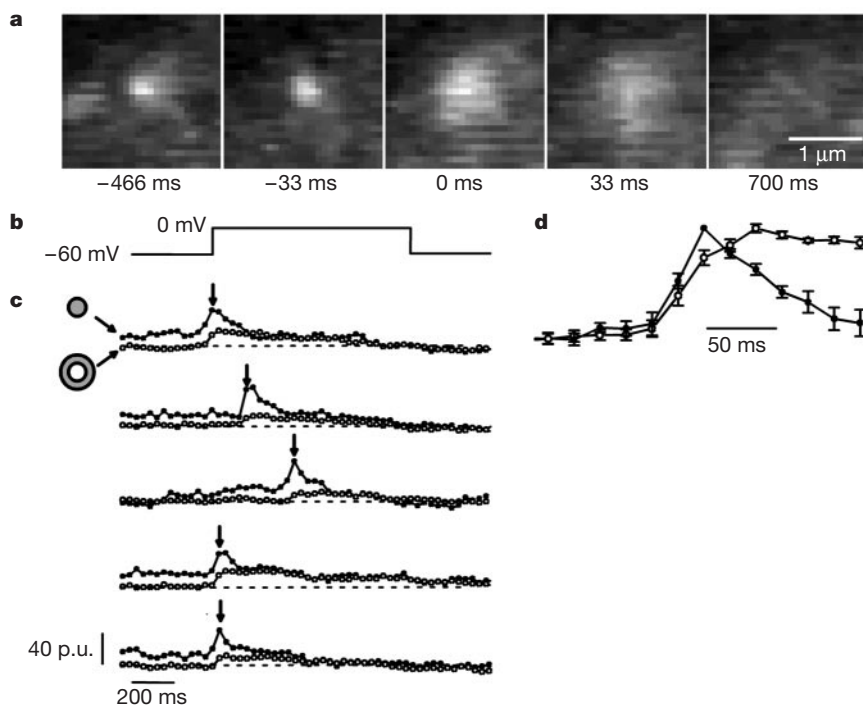


Figure 2 Imaging single fusion events. **a**, Images of a vesicle fusing with the plasma membrane. Times of recording are relative to the beginning of the voltage step in **b**. **c**, Fluorescence intensities were plotted for small circles enclosing fluorescent spots (filled symbols) and for concentric annuli around the circles (open symbols). Arrows mark fusion,

ordinate in pixel units (p.u.) of brightness. **d**, Average of 10 traces including those in **c**, normalized to maximum brightness and aligned to time of fusion. In **d**, frames were deinterlaced before analysis to double the time resolution. All data are from the same cell.

pulse. Before a pulse or without it, fusions were rare (0.024 s^{-1} ; 5 events in 211 s at -60 mV in 16 cells). During the first 67 ms of a pulse, the fusion rate rose 200-fold and then declined. The actual peak rate was undoubtedly higher, as the most rapidly releasable vesicle pool in bipolar terminals is exhausted with time constants of milliseconds^{6,9}. Exocytosis outlasted the pulse, perhaps because internal $[\text{Ca}^{2+}]$ remained elevated (fusion rate 0.13 s^{-1} during the first 0.5 s after a pulse, 76 s of recording in 16 cells). Evidently the $\sim 1 \mu\text{M}$ $[\text{Ca}^{2+}]$ expected after a pulse^{25,26} stimulates fusion of some vesicles^{20,21}.

Like previous work using other methods⁶⁻⁹, Fig. 4a shows a fast and a slower phase, suggesting fast and slow vesicle pools. To test whether vesicles are slow because they have yet to 'dock' at the plasma membrane, we used the fact that the evanescent field declines exponentially with distance from the glass-cell interface. Illuminated in proportion to their proximity to the interface, fluorescent vesicles will brighten as they approach and dim as they retreat from the plasma membrane. Apart from fusion, approaches and retreats are the most likely explanation for fluorescence changes, as the fluorescence of membrane-bound FM1-43 does not depend strongly on either the membrane potential^{9,27} or

the pH (ref. 28 and data not shown). We distinguished two classes of vesicles. 'Residents' were visible during the entire 0.5 s before the voltage step, did not increase in brightness until they fused (Fig. 4d) and are equivalent to morphologically docked vesicles (see Methods). 'Newcomers' were absent or only dimly visible 0.5 s before the step and brightened before fusion (Fig. 4e, f). The rapid phase of exocytosis was largely due to residents (Fig. 4b) and the slow phase to newcomers (Fig. 4c). Evidently the fastest pool of vesicles detected in our assay largely coincides with that of docked vesicles, and further exocytosis requires capture of new vesicles.

Figure 5a shows the approach of a newcomer to the plasma membrane. Fluorescence rose to a nearly constant value (Fig. 5b), indicating the capture of a vesicle at a site on or near the plasma membrane (dotted line). Similar trajectories from other vesicles were aligned to the time of approach and averaged. The result can be modelled (see movie 4 in Supplementary Information) by assuming the vesicle to approach at a constant speed of $0.8 \mu\text{m s}^{-1}$ and to stop abruptly at its destination (Fig. 5c). In Fig. 5d a vesicle approached, remained for about 0.5 s and then vanished. Unlike in Fig. 2, its fluorescence never spread, so it must have detached rather than fused. Evidently capture is reversible. Figure 5f seems to indicate that some vesicles detached from capture sites with a rate constant of more than 0.25 s^{-1} .

How soon after fusion are residents replaced by newcomers? Figure 6a plots the number of fusions and arriving newcomers against time. The voltage step stimulated the capture of new vesicles near the plasma membrane. Within 2 s, most fusing vesicles were replaced by newcomers. To calculate how quickly newcomers reach exocytic competence, Fig. 6b shows the time course whereby newcomers are lost through exocytosis. All vesicles arriving within the

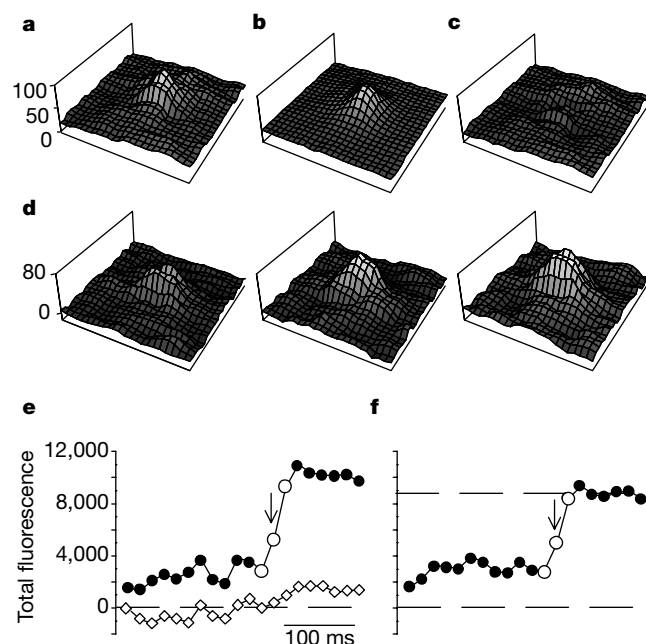


Figure 3 Size of the dye-containing organelle. **a**, Distribution of fluorescence in a $2.6 \times 2.6 \mu\text{m}$ square centred on a vesicle imaged 33–17 ms before fusion. **b**, Point-spread function (PSF) superimposed on an inclined plane. The PSF amplitude, its width at half maximum, the x - y coordinates of its centre and the constants defining the plane were adjusted to provide the best least-squares fit to **a**. The PSF was the image of a 40-nm diameter fluorescent bead. **c**, Background image (difference between **a** and the fitted PSF in **b**). **d**, Consecutive images shortly after fusion, with the image in **c** subtracted. **e**, Each circle sums the light over all pixels of a $2.6 \times 2.6 \mu\text{m}$ square, with background subtracted as in **d**. Summed light from consecutive frames plotted against time; arrow indicates fusion. Open circles from images in **d**. Diamonds indicate estimated time-dependent gain of dye from fusion events elsewhere in the cell. They track light from regions of the terminal's footprint devoid of fusion events or bright spots. Light was divided by $a/6.86 \mu\text{m}^2$ where a is the area of the region and $6.86 \mu\text{m}^2$ that of the $2.6 \times 2.6 \mu\text{m}$ square. **f**, Difference between the two curves in **e** tracks light from the dye originally contained in the vesicle. Fusion causes brightening; the plateau (horizontal line) measures the light (9,572 units for this vesicle) after all dye has been released into the plasma membrane. Similar analysis was done on 17 other fusions from the cell in Fig. 1, chosen because they left no trace of the vesicle and hence resulted in complete release of dye, occurred $> 1.2 \mu\text{m}$ from the edge of the cell, and did not overlap with other fusions occurring at about the same time. Mean plateau value, $7,656 \pm 413$ units. De-interlaced frames throughout analysed in Mathematica (Wolfram Research).

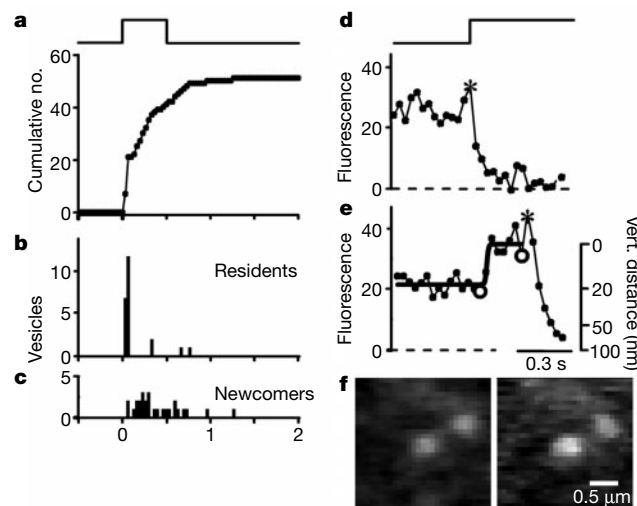


Figure 4 Kinetics of exocytosis. **a**, Voltage (top) and cumulative number of fusion events (bottom) plotted against time since stimulus onset. Times of fusion determined as in Fig. 2c. **b**, **c**, Times of fusion plotted as histograms for residents (**b**) and newcomers (**c**) distinguished by the same-vesicle test. Data from 13 cells with mean Ca^{2+} current of $116 \pm 5 \text{ pA}$. **d**, Fluorescence from a resident vesicle plotted against time relative to the stimulus (above). Intensity in a 524-nm diameter circle after subtraction of background (annulus with 524 nm inner and 1,571 nm outer diameters). Asterisk, time of fusion as in **e**. **e**, Fluorescence, I , of a newcomer visible from the beginning of the recording period, measured as in **d**. Curve is a modified Boltzmann distribution, $I = A + B/[1 + \exp[(t_0 - t)/T]]$ where t is time and A , B , t_0 and T are constants adjusted to provide the best least-squares fit. The curve was used to estimate objectively the timing of the vesicle's approach in a noisy trace. Right ordinate, vesicle's distance d from plasma membrane calculated from I . The average I during the last three frames before fusion, I_0 , was assumed to represent the brightness of the vesicle while it touched the plasma membrane ($d = 0$). Other values of d calculated by $d = D \ln(I_0/I)$ where $D = 42 \text{ nm}$ is the distance over which the evanescent field declines e-fold. **f**, Images of vesicle in **e** taken at the times indicated by open circles.

first 270 ms of a pulse fused with a mean latency of 250 ± 30 ms ($n = 7$), comparable to an earlier estimate²⁹. In a larger data set including all newcomers fusing during a pulse, the mean latency was 210 ± 30 ms ($n = 18$). The latency estimates the time required for 'priming', defined here as the time between docking and exocytic competence. In priming, vesicles must complete the assembly of the entire macromolecular complex needed to sense $[Ca^{2+}]$ and then catalyse membrane fusion. Together with the recruitment of new vesicles, the need for priming may contribute to the interval sometimes seen to separate early and late components of exocytosis^{7,8}.

Properties of active zones

Preferred sites of vesicle residence (Fig. 1b) suggest preferred sites of exocytosis, or active zones. This was confirmed when fusion sites were marked on time-averaged images (red dots in Fig. 1d). Of the 40 fusions in Fig. 1d, 29 occurred at 5 bright spots where vesicles resided in preference. Figure 1d also shows spatially isolated vesicles

fusing at sites other than bright spots. Similar results were seen in all six cells where we recorded more than 13 events. We found that $64 \pm 6\%$ of fusions occurred within 300 nm of another fusion, and hence at an active zone. Active zones and bright spots largely coincided, so we view both as manifestations of the same specialization within or beneath the plasma membrane. The number of active zones was 0.10 ± 0.03 per μm^2 of plasma membrane ($n = 7$ cells) and comparable to that of synaptic ribbons ($0.20 \mu m^{-2}$, calculated from ref. 19).

Vesicles at bright spots and active zones approached the plasma membrane no faster than they did elsewhere (open and filled circles in Fig. 5c), nor were they more easily captured. Once resident, however, they fused more readily. At bright spots in time-averaged images, $90 \pm 9\%$ of resident vesicles fused during a 0.5-s voltage step, compared with $43 \pm 9\%$ elsewhere (four cells, $P < 0.005$). Fusions also occurred earlier at active zones. When vesicles fused elsewhere, their chance of fusing in the first 67 ms after a step was only 0.56 ± 0.15 times that at active zones ($n = 6$ cells, $P < 0.005$), largely because most fusing vesicles elsewhere were newcomers.

Some newcomers (28% in 5 cells) appeared to be stationary and were dimly visible for more than 0.33 s before making a final approach (see, for example, Fig. 4e, f). This was more often true at active zones (12 of 27; $P < 0.02$) than elsewhere (3 of 29; $P < 0.05$). Evidently structures exist at active zones that hold vesicles captive a short distance beneath the plasma membrane and, during stimulation, allow their final approach to the plasma membrane. When such vesicles fused (Fig. 4e, f, and movie 5 in Supplementary Information) it was possible to estimate how far from the plasma membrane they were held before their final approach. The average distance was 20 ± 4 nm ($n = 6$ vesicles in 4 cells).

Discussion

Synaptic terminals in retinal bipolar cells contain ribbons, sub-micrometre-sized electron-dense structures beneath the plasma

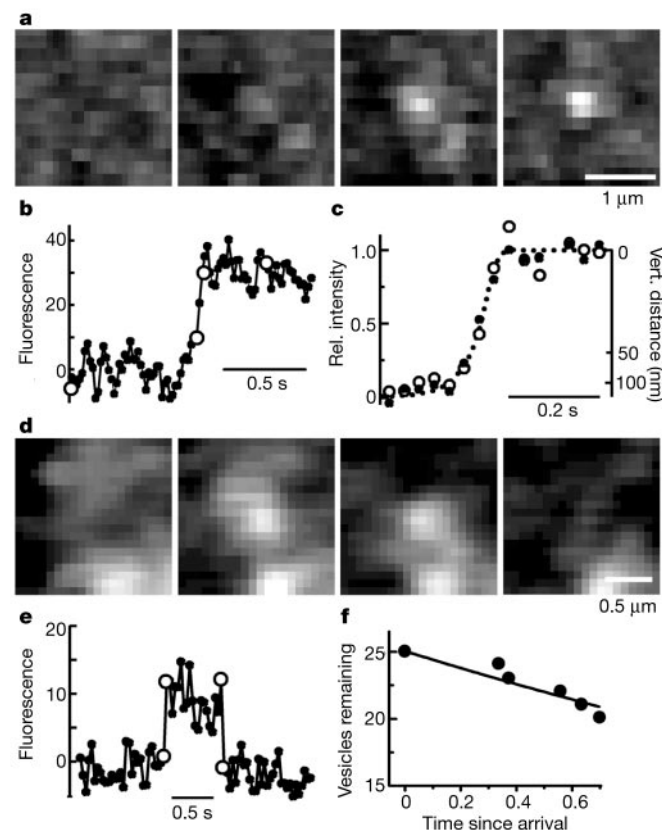


Figure 5 Capture and escape of vesicles. **a**, Sequential images of an approaching vesicle. **b**, Fluorescence at the vesicle site; open circles from the four images in **a**. **c**, 22 trajectories as in **b** were aligned to their midpoint where the fluorescence has risen to half its final value, and averaged. Midpoint determined by fitting a Boltzmann curve as in Fig. 4e. We excluded vesicles that were visible at the beginning of the trace (for example, Fig. 4e, f) but accepted vesicles approaching both at bright spots as in Fig. 1b (open circles, 6 vesicles) and elsewhere (filled circles, 17 vesicles). Similar results were obtained at active zones and elsewhere. Right ordinate, distance from destination calculated as in Fig. 4e. A model calculation (dotted line) assumed newcomers to approach at constant velocity and to stop abruptly at their destination; it took account of the 33.3-ms integration time required to form an image. Approach velocity of $0.8 \mu m s^{-1}$ provides the best least-squares fit to the filled circles. **d**, Sequential images taken at the times indicated by open circles in **e**; note that the vesicle at the centre is absent in the first and last frames. Images low-pass filtered for clarity. **e**, Fluorescence at the vesicle site. **f**, Vesicles arriving > 0.8 s before the end of the recording period were selected and their fluorescence traces aligned to arrival time. Of 25 vesicles, 5 detached during the 0.8-s interval. Continuous line, declining exponential with time constant 4 s, corresponding to escape rate constant of $0.25 s^{-1}$.

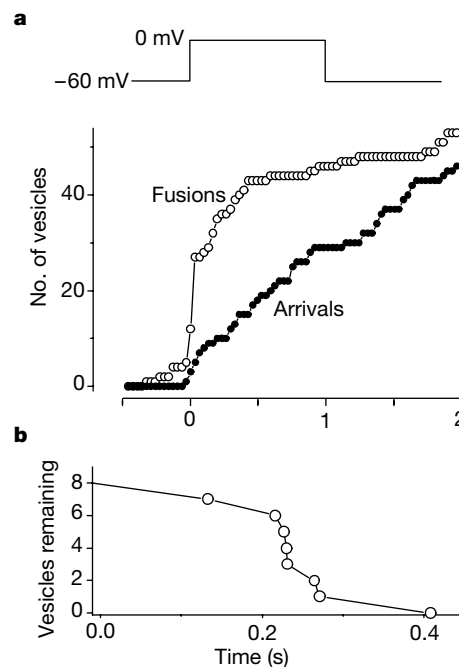


Figure 6 Stimulated approach of vesicles. **a**, Cumulative number of fusions (open) and arrivals (filled) against time in seconds. Fusion times determined as in Fig. 2c. Arrival times of newcomers determined from Boltzmann curves fitted as in Fig. 4e. Newcomers were considered to have 'arrived' (finished their approach) when the Boltzmann curve had reached 90% of its final value. **b**, Loss through exocytosis of newcomers arriving within the first 200 ms of a voltage pulse; all fused during the pulse while Ca^{2+} channels remained open.

membrane. Morphologists believe that they transport vesicles to the plasma membrane³⁰, capture and position them against the plasma membrane³¹ and mark active zones. Our results show the existence of active zones of exocytosis that largely coincide with small bright spots where vesicles are captured and reside in preference. We suggest that bright spots and active zones mark ribbons. Solitary vesicles also fused and were captured at scattered sites elsewhere on the plasma membrane. As we could not actually see ribbons, we cannot exclude the possibility that each scattered site represents an additional ribbon, albeit of diminished activity. An attractive alternative is suggested by findings in hair cells of the inner ear, where presynaptic terminals contain dense bodies closely analogous to ribbons. In hair cells, morphologically docked vesicles not only cluster around dense bodies but also populate remote sites as solitary 'outliers'³². If some solitary vesicles in our work were outliers, it follows that ribbons are not absolutely required for docking and exocytosis. Nor do they help in transport, as vesicles are transported equally rapidly to all parts of the plasma membrane. However, ribbons do provide sites where vesicles fuse earlier than elsewhere, and where docked vesicles fuse at high probability, possibly because calcium channels congregate there³³.

Most importantly, ribbons may provide the structure holding vesicles 'in reserve' at a distance of about 20 nm until a stimulus triggers movement to the plasma membrane. This distance is about twice the cytosolic length of a VAMP/synaptobrevin molecule, or of the coiled bundle of v- and t-SNAREs thought to form during exocytosis (length 12 nm (ref. 34)). Hence vesicles at 20 nm from the plasma membrane are close enough for v- and t-SNAREs to touch and, possibly, to initiate formation of the coiled complex. Indeed, SNARE pairing may contribute to the last 20 nm of vesicle movement. Ribbons may facilitate SNARE complex formation, and thus be functionally analogous to the p115/GM130 complex tethering vesicles to Golgi cisternae as a prelude to fusion³⁵. □

Methods

Cells, labelling and electrical recording

Goldfish retinae³⁶ were plated on coverslips of high refractive index glass ($n = 1.80$ at 488 nm) and kept until use in low- Ca^{2+} bath solution (containing, in mM, 0.5 CaCl_2 , 120 NaCl, 2.5 KCl, 1.0 MgCl_2 , 10 glucose, 10 Na-HEPES, pH 7.4). Bipolar cells were recognized by their bulbous synaptic terminals at the ends of short axons, and selected under reflection interference contrast optics for tight adherence of the terminal to the coverslip. Selected cells were loaded with FM1-43 (Molecular Probes) by local superfusion. Cells were superfused first with a solution containing 5 μM FM1-43 plus (in mM) 2.5 CaCl_2 , 95 NaCl, 25 KCl, 1.0 MgCl_2 , 10 glucose, 10 Na-HEPES, pH 7.4, and then for 15–30 min with dye-free low- Ca^{2+} bath solution with 1.25 mM EGTA added to further lower $[\text{Ca}^{2+}]$ to about 100 nM Ca^{2+} . The procedure is expected to label about 1% of all vesicles inside a terminal²¹. Simultaneously a peristaltic pump perfused the whole chamber with EGTA-containing low- Ca^{2+} solution. After staining, cells were kept in a Ca^{2+} -containing bath solution (in mM, 2.5 Ca^{2+} , 120 NaCl, 2.5 KCl, 1.0 MgCl_2 , 10 glucose, 10 Na-HEPES, pH 7.4). We voltage clamped the somata in the whole-cell configuration and held them at -60 mV potential with an EPC-9 amplifier (Heka Elektronik). The pipette solution contained (in mM): 120 Cs^+ glutamate, 4 Na_2ATP , 0.5 GTP, 4 MgCl_2 , 10 tetraethylammonium-Cl, 0.5 EGTA and 10 Cs-HEPES (pH 7.4). Exocytosis was stimulated by changing the potential to 0 mV for 0.5 or 1 s. Rest intervals between stimuli were 30–60 s. Data obtained > 8 min after patch break were discarded to limit the effects of run down and photobleaching. Experiments were done at room temperature, means are given $\pm$ standard error.

Optical imaging

An inverted microscope was modified for evanescent field epi-illumination^{14,37,38}. An expanded laser beam (488 nm) was focused at the back focal plane of a high numerical aperture objective lens (Planapo $\times 100$, 1.65 NA, Olympus), and the focal point moved off-axis to the most peripheral position that still allowed the beam to pass through the objective. P-polarized excitation light was used³⁹. An immersion oil (sulphur in diiodomethane, refractive index $n = 1.81$ at 488 nm, Cargille Laboratories) made optical contact between the objective and the coverslip. Light entered the coverslip at an angle measured as 65.6 – 68.0° and underwent total internal reflection at the glass–cell interface. Scattered and back-reflected excitation light was blocked from the emission path by a dichroic mirror (505DRLP, Omega Optical) and a holographic notch filter (HSPF-488.0-1.0, Kaiser Optical Systems). The refractive indices for glass ($n = 1.80$ at 488 nm) and cell ($n = 1.37$) predict an evanescent field declining e-fold within 41–43 nm from the interface, and to $\sim 10\%$ within 100 nm. A vesicle 100 nm from the interface would be illuminated too dimly to be visible under our conditions. Hence we look barely 100 nm into the cell, a

distance comparable to the thickness of ultrathin sections in electron microscopy. Images were intensified (VS4-1845, Video Scope International), recorded at 30 interlaced frames per s by a video camera (VarioTrig, PCO Computer Optics) and stored on optical disc (TQ3038F, Panasonic). One pixel represented 87.3 nm at the cell surface. Voltage pulses and image capture were synchronized to < 2 video frames (67 ms).

Analysis

Video images were digitized and analysed using MetaMorph 4.0 (Universal Imaging). Unless otherwise indicated, fluorescence of single vesicles was measured as the brightness difference between a 873 nm diameter circle centred on the vesicle, and a concentric annulus with 873 nm inner and 2,182 nm outer diameter. Brightness changes of individual vesicles were interpreted as fusions if the fluorescence increased, spread and then declined. Otherwise they were taken to indicate movements towards or away from the plasma membrane, with an e-fold brightness change representing a 42-nm movement. The distance moved is actually less because not only the illumination but also the efficiency of light collection diminishes with distance from the interface³⁹. This effect is ignored here.

Two specialized regions were defined on the surface of a terminal. The first was an 870×870 nm square centred on a bright spot in a time-averaged image as in Fig. 1b, d. The second, or active zone, included any location within 300 nm of a fusion site located < 300 nm away from another.

We used two tests to decide whether vesicles were resident (or docked) at the plasma membrane. The 'same-vesicle' test can be used only on fusing vesicles and requires that the brightness of a vesicle 0.4–0.5 s before a voltage pulse be $> 80\%$ of that immediately before fusion (average $102 \pm 2\%$, 60 vesicles in 11 cells). In an evanescent field declining e-fold in 42 nm, a vesicle at 80% brightness would have a vertical distance of 9.3 nm from its fusion site at the plasma membrane and qualify as morphologically docked (vesicle distance from plasma membrane < 10 nm in electron microscopic studies^{32,40}). The less reliable 'same-cell' test can also be applied to vesicles that do not fuse. It accords residence status based on the pre-stimulus brightness of a vesicle, defined as the average brightness during the penultimate 14 video frames before a voltage step. For each cell, we first found all fusing vesicles that passed the same-vesicle test for residence. We then determined their pre-stimulus brightness and used the mean value as a benchmark. Vesicles in that cell were considered resident if their pre-stimulus brightness was $> 80\%$ of the benchmark.

Received 27 April; accepted 20 July 2000.

- Liley, A. W. & North, K. A. K. An electrical investigation of effects of a repetitive stimulation on mammalian neuromuscular junction. *J. Neurophysiol.* **16**, 509–527 (1953).
- Zucker, R. S. Exocytosis: a molecular and physiological perspective. *Neuron* **17**, 1049–1055 (1996).
- Neher, E. Vesicle pools and Ca^{2+} microdomains: new tools for understanding their roles in neurotransmitter release. *Neuron* **20**, 389–399 (1998).
- Wu, L. G. & Borst, J. G. The reduced release probability of releasable vesicles during recovery from short-term synaptic depression. *Neuron* **23**, 821–832 (1999).
- Schneggenburger, R., Meyer, A. C. & Neher, E. Released fraction and total size of a pool of immediately available transmitter quanta at a calyx synapse. *Neuron* **23**, 399–409 (1999).
- Mennerick, S. & Matthews, G. Ultrafast exocytosis elicited by calcium current in synaptic terminals of retinal bipolar neurons. *Neuron* **17**, 1241–1249 (1996).
- Sakaba, T., Tachibana, M., Matsui, K. & Minami, N. Two components of transmitter release in retinal bipolar cells: exocytosis and mobilization of synaptic vesicles. *Neurosci. Res.* **27**, 357–370 (1997).
- von Gersdorff, H., Sakaba, T., Berglund, K. & Tachibana, M. Submillisecond kinetics of glutamate release from a sensory synapse. *Neuron* **21**, 1177–1188 (1998).
- Neves, G. & Lagnado, L. The kinetics of exocytosis and endocytosis in the synaptic terminal of goldfish retinal bipolar cells. *J. Physiol. (Lond.)* **515**, 181–202 (1999).
- Steyer, J. A., Horstmann, H. & Almers, W. Transport, docking and exocytosis of single secretory granules in live chromaffin cells. *Nature* **388**, 474–478 (1997).
- Oheim, M., Loecker, D., Stuhmer, W. & Chow, R. H. The last few milliseconds in the life of a secretory granule. Docking, dynamics and fusion visualized by total internal reflection fluorescence microscopy (TIRFM). *Eur. Biophys. J.* **27**, 83–98 (1998).
- Han, W., Ng, Y. K., Axelrod, D. & Levitan, A. E. Neuropeptide release by efficient recruitment of diffusing cytoplasmic secretory vesicles. *Proc. Natl Acad. Sci. USA* **96**, 14577–14582 (1999).
- Lang, T. et al. Role of actin cortex in the subplasmalemmal transport of secretory granules in PC-12 cells. *Biophys. J.* **78**, 2863–2877 (2000).
- Schmoranzner, J., Goulian, M., Axelrod, D. & Simon, S. M. Imaging constitutive exocytosis with total internal reflection fluorescence microscopy. *J. Cell Biol.* **149**, 23–32 (2000).
- Toomre, D., Steyer, J. A., Keller, P., Almers, W. & Simons, K. Fusion of constitutive membrane traffic with the cell surface observed by evanescent wave microscopy. *J. Cell Biol.* **149**, 33–40 (2000).
- Betz, W. J. & Bewick, G. S. Optical analysis of synaptic vesicle recycling at the frog neuromuscular junction. *Science* **255**, 200–203 (1992).
- Ryan, T. A., Reuter, H. & Smith, S. J. Optical detection of a quantal presynaptic membrane turnover. *Nature* **388**, 478–482 (1997).
- Murthy, V. N. & Stevens, C. F. Synaptic vesicles retain their identity through the endocytic cycle. *Nature* **392**, 497–501 (1998).
- von Gersdorff, H., Vardi, E., Matthews, G. & Sterling, P. Evidence that vesicles on the synaptic ribbon of retinal bipolar neurons can be rapidly released. *Neuron* **16**, 1221–1227 (1996).
- Lagnado, L., Gomis, A. & Job, C. Continuous vesicle cycling in the synaptic terminal of retinal bipolar cells. *Neuron* **17**, 957–967 (1996).
- Rouze, N. C. & Schwartz, E. A. Continuous and transient vesicle cycling at a ribbon synapse. *J. Neurosci.* **18**, 8614–8624 (1998).
- Takei, K., Mundigl, O., Daniell, L. & De Camilli, P. The synaptic vesicle cycle: a single vesicle budding step involving clathrin and dynamin. *J. Cell Biol.* **133**, 1237–1250 (1996).
- Schmidt, A., Hannah, M. J. & Huttner, W. B. Synaptic-like microvesicles of neuroendocrine cells originate from a novel compartment that is continuous with the plasma membrane and devoid of transferrin receptor. *J. Cell Biol.* **137**, 445–458 (1997).
- Zhang, B. et al. Synaptic vesicle size and number are regulated by a clathrin adaptor protein required for endocytosis. *Neuron* **21**, 1465–1475 (1998).

25. von Gersdorff, H. & Matthews, G. Inhibition of endocytosis by elevated internal calcium in a synaptic terminal. *Nature* **370**, 652–655 (1994).
26. Kobayashi, K. & Tachibana, M. Ca^{2+} regulation in the presynaptic terminals of goldfish retinal bipolar cells. *J. Physiol. (Lond.)* **483**, 79–94 (1995).
27. Smith, C. B. & Betz, W. J. Simultaneous independent measurement of endocytosis and exocytosis. *Nature* **380**, 531–534 (1996).
28. Betz, W. J., Mao, F. & Smith, C. B. Imaging exocytosis and endocytosis. *Curr. Opin. Neurobiol.* **6**, 365–371 (1996).
29. Betz, W. J. & Bewick, G. S. Optical monitoring of transmitter release and synaptic vesicle recycling at the frog neuromuscular junction. *J. Physiol. (Lond.)* **460**, 287–309 (1993).
30. Gray, E. G. & Pease, H. L. On understanding the organisation of the retinal receptor synapses. *Brain Res.* **35**, 1–15 (1971).
31. Raviola, E. & Gilula, N. B. Intramembrane organization of specialized contacts in the outer plexiform layer of the retina. A freeze-fracture study in monkeys and rabbits. *J. Cell Biol.* **65**, 192–222 (1975).
32. Lenzi, D., Runyeon, J. W., Crum, J., Ellisman, M. H. & Roberts, W. M. Synaptic vesicle populations in saccular hair cells reconstructed by electron tomography. *J. Neurosci.* **19**, 119–132 (1999).
33. Issa, N. P. & Hudspeth, A. J. Clustering of Ca^{2+} channels and Ca^{2+} -activated K^{+} channels at fluorescently labeled presynaptic active zones of hair cells. *Proc. Natl Acad. Sci. USA* **91**, 7578–7582 (1994).
34. Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395**, 347–353 (1998).
35. Lowe, M., Nakamura, N. & Warren, G. Golgi division and membrane traffic. *Trends Cell Biol.* **8**, 40–44 (1998).
36. Heidelberger, R. & Matthews, G. Calcium influx and calcium current in single synaptic terminals of goldfish retinal bipolar neurons. *J. Physiol. (Lond.)* **447**, 235–256 (1992).
37. Stout, A. L. & Axelrod, D. Evanescent field excitation of fluorescence by epi-illumination microscopy. *Appl. Opt.* **28**, 5237–5242 (1989).
38. Steyer, J. A. & Almers, W. Tracking single secretory granules in live chromaffin cells by evanescent-field fluorescence microscopy. *Biophys. J.* **76**, 2262–2271 (1999).
39. Axelrod, D., Hellen, E. H. & Fulbright, R. M. in *Topics in Fluorescence Spectroscopy, Volume 3: Biomedical Applications* (ed. Lakowicz, J. R.) 289–343 (Plenum, New York, 1992).
40. Parsons, T. D., Coorssen, J. R., Horstmann, H. & Almers, W. Docked granules, the exocytic burst, and the need for ATP hydrolysis in endocrine cells. *Neuron* **15**, 1085–1096 (1995).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>).

Acknowledgements

This work was supported by a grant from the NIH (W.A.) and by the Max Planck Society. D.Z. was supported by an NIH Training Program in Neuronal Signaling and J.A.S. by a fellowship from the Max Planck Society. We thank M. Feldman for his comments on the manuscript.

Correspondence and requests for materials should be addressed to W.A. (e-mail: almersw@ohsu.edu).

Structural and biochemical basis of apoptotic activation by Smac/DIABLO

Jijie Chai*, Chunying Du†, Jia-Wei Wu*, Saw Kyin*, Xiaodong Wang† & Yigong Shi*

* Department of Molecular Biology, Princeton University, Washington Road, Princeton, New Jersey 08544, USA

† Howard Hughes Medical Institute and Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

Apoptosis (programmed cell death), an essential process in the development and homeostasis of metazoans, is carried out by caspases. The mitochondrial protein Smac/DIABLO performs a critical function in apoptosis by eliminating the inhibitory effect of IAPs (inhibitor of apoptosis proteins) on caspases. Here we show that Smac/DIABLO promotes not only the proteolytic activation of procaspase-3 but also the enzymatic activity of mature caspase-3, both of which depend upon its ability to interact physically with IAPs. The crystal structure of Smac/DIABLO at 2.2 Å resolution reveals that it homodimerizes through an extensive hydrophobic interface. Missense mutations inactivating this dimeric interface significantly compromise the function of Smac/DIABLO. As in the *Drosophila* proteins Reaper, Grim and Hid, the amino-terminal amino acids of Smac/DIABLO are indispensable for its function, and a seven-residue peptide derived from the amino terminus promotes procaspase-3 activation *in vitro*. These results establish an evolutionarily conserved structural and biochemical basis for the activation of apoptosis by Smac/DIABLO.

Apoptosis is crucial in the development and homeostasis of all multicellular organisms^{1–4}. Abnormal inhibition of apoptosis is a hallmark of cancer and autoimmune diseases, whereas excessive cell death is implicated in neurodegenerative disorders such as Alzheimer disease^{5,6}. The mechanism of apoptosis is remarkably conserved across species, involving a cascade of initiator and effector caspases that are activated sequentially^{7,8}.

Caspases, a family of cysteine proteases with aspartate substrate specificity, are produced in cells as catalytically inactive zymogens⁷. Effector caspases, such as caspase-3, are activated by initiator caspases, such as caspase-9, through proteolytic cleavage at specific internal Asp residues⁷. Once activated, the effector caspases are responsible for proteolytic cleavage of a range of cellular targets, ultimately leading to cell death.

The inhibitor of apoptosis (IAP) family of proteins, originally identified in the genome of baculovirus on the basis of their ability to suppress apoptosis in infected host cells, has a key function in the negative regulation of programmed cell death in a variety of organisms^{9,10}. IAPs suppress apoptosis by preventing the activation of procaspases and inhibiting the enzymatic activity of mature

caspases^{9,10}. Several distinct mammalian IAPs including XIAP, c-IAP1, c-IAP2 and survivin have been identified, and they all exhibit anti-apoptotic activity in cell culture^{9,10}. In *Drosophila*, the anti-apoptotic activity of IAPs is removed by Reaper, Grim and Hid, all of which appear to act upstream of IAPs and interact physically with IAPs to relieve their inhibitory effect on caspase activation^{11,12}.

One important caspase activation cascade is triggered by the release of cytochrome *c* from the intermembrane space of mitochondria, which occurs in response to several apoptotic stimuli including serum deprivation, DNA damage and activation of cell-surface death receptors^{13,14}. In the cytosol, cytochrome *c* associates with Apaf-1 in the presence of dATP or ATP and induces its oligomerization. The oligomeric Apaf-1 complex recognizes the inactive procaspase-9, forming the 'apoptosome', which induces autocatalytic processing of procaspase-9 (refs 15–19). The mature caspase-9 in turn activates its primary downstream target procaspase-3.

Concurrent with cytochrome *c* release, another important regulator of apoptosis, Smac²⁰ (second mitochondria-derived activator of caspases) or DIABLO²¹, is also released from the mitochon-

Table 1 Summary of crystallographic analysis

	Resolution (Å)	Reflections		Overall (outer shell)			MIR analysis (20–2.1 Å)		
		Measured	Unique	Data coverage (%)	R_{sym} (%)	Heavy atom sites	All (outer shell) isomorphous difference (%)	Acentric/centric Phasing power	Cullis R factor
Native 1	2.2	97,019	12,842	99.5 (97.9)	4.7 (24.2)				
EMP	2.9	23,399	4,671	80.7 (81.5)	9.9 (33.6)	3	23.7 (25.6)	1.85/1.28	0.67/0.65
HgAc	3.5	14,640	3,037	90.3 (92.4)	10.0 (30.2)	2	21.5 (23.7)	1.43/0.78	0.75/0.81
K ₃ UO ₃ F ₅	3.6	9,213	2,630	86.0 (88.2)	10.2 (26.7)	4	29.5 (28.5)	1.11/0.71	0.85/0.82
PhNH ₂ HgAc	3.5	9,710	3,132	94.5 (95.0)	9.5 (23.8)	4	21.8 (25.7)	1.35/0.85	0.77/0.77
MeHgAc	3.5	7,738	2,925	87.7 (92.3)	10.2 (28.5)	4	19.5 (31.7)	0.90/0.53	0.85/0.88
Se-Met	3.4	20,598	6,057	95.9 (95.5)	10.2 (24.9)	4	16.0 (22.0)	1.12/0.75	0.82/0.78
Thimerosal	3.2	18,133	4,285	98.5 (99.0)	10.7 (33.3)	4	16.0 (30.4)	1.46/0.94	0.73/0.75
Mean figure of merit (20.0–3.0 Å): 0.586									
Refinement	Resolution range	Reflections ($F > 2\sigma$)	Atoms modelled (total, water)	$R_{\text{working}}/R_{\text{free}}$ (%)	r.m.s. deviation				
					Bonds (Å)	Angles (deg)	B factor (Å ²)		
	15.0–2.2	11,728	1,420,47	24.5/25.9	0.006	1.0120	3.880		

Se-Met: seleno-methionine-derivatized Smac. $R_{\text{sym}} = \sum_i \sum_j |I_{ij} - I_j| / \sum_j I_j$, where I_j is the mean intensity of the j observations of symmetry related reflections of h . Isomorphous difference = $\sum_i |F_{\text{PH}} - F_{\text{P}}| / \sum_i |F_{\text{PH}}|$, where F_{P} and F_{PH} are the native and derivative structure factor amplitudes, respectively. Phasing power = $[(F_{\text{H(calc)}})^2 / (F_{\text{PH(obs)}})^2 - (F_{\text{H(calc)}})^2]^{1/2}$, where $F_{\text{PH(obs)}}$ and $F_{\text{H(calc)}}$ are the observed and calculated derivative structure factors, respectively. Cullis R factor = $\sum_i |F_{\text{PH}} \pm F_{\text{P}}| - F_{\text{H(calc)}} / \sum_i |F_{\text{PH}} \pm F_{\text{P}}|$, where $F_{\text{H(calc)}}$ is the calculated heavy atom structure factor. Figure of merit = $\langle \Sigma P(\alpha) \exp(i\alpha) / \Sigma P(\alpha) \rangle$, where $P(\alpha)$ is the probability distribution for the phase α . $R = \sum_i |F_{\text{obs}} - F_{\text{calc}}| / \sum_i |F_{\text{obs}}|$, where $F_{\text{obs}} = F_{\text{P}}$, and F_{calc} is the calculated protein structure factor from the atomic model (R_{free} was calculated with 5% of the reflections). r.m.s. in bond lengths and angles are the deviations from ideal values, and the r.m.s. deviation in B factors is calculated between bonded atoms.

dria into the cytosol. Whereas cytochrome *c* induces multimerization of Apaf-1 to activate procaspase-9 and -3, Smac eliminates the inhibitory effect of many IAPs^{20,21}. Smac interacts with all IAPs that have been examined, including XIAP, c-IAP1, c-IAP2 and survivin^{20,21}. Thus, Smac appears to be a master regulator of apoptosis in mammals and a functional homologue of the *Drosophila* proteins Reaper, Grim and Hid.

Smac is synthesized as a 239-amino-acid precursor molecule; the amino-terminal 55 residues serve as the mitochondria targeting sequence, which is removed after import²⁰. The mature form of Smac contains 184 amino acids and behaves as an oligomer in solution²⁰. Despite its importance in cell death, no structural information is available on Smac or on its *Drosophila* functional homologues Reaper, Grim and Hid.

Here, we report the 2.2 Å resolution crystal structure of mature Smac, which shows an arch-shaped homodimer with rich surface features. The homodimeric interface is dominated by hydrophobic residues through van der Waals interactions. Mutations of key residues at the interface disrupt dimer formation and significantly weaken the ability of Smac to activate procaspase-3 and promote the enzymatic activity of mature caspase-3. In addition, the N-terminal residues of mature Smac are essential for Smac function, as mutation of the first amino acid renders the resulting protein completely inactive. Furthermore, we show that the N-terminal peptides of Smac can promote procaspase-3 activation *in vitro*, suggesting therapeutic potential. Combining structural, mutational and bio-

chemical analyses, we propose a coherent model for apoptotic activation by Smac.

Structure of a Smac monomer

The mature form of Smac (residues 1–184; relative molecular mass 21,000 (M_r 21K)) was overexpressed in bacteria, purified to homogeneity and crystallized. The X-ray structure was determined at 2.2 Å resolution by multiple isomorphous replacement (Table 1). Our final atomic model contains residues 11–182. No electron density is seen corresponding to the N-terminal 10 residues, and we presume that this region is disordered in solution.

The Smac monomer is an elongated three-helix bundle with moderate curvature (Fig. 1a). The H1 helix packs closely against helices H2 and H3, whereas H2 and H3 are farther apart. Most of the hydrophobic residues are located in the interior of the three-helix bundle, contributing to its stability. In addition, the elongated helices are buttressed by 23 intra- and 9 interhelical hydrogen bonds involving the side chains of exposed polar residues. Owing to its elongated shape, Smac is expected to exhibit a much larger radius of hydration in solution than a globular protein with the same molecular mass.

Around half of the exposed hydrophobic residues cluster on one small surface patch formed by the N-terminal half of H1 and the carboxy-terminal third of H2 (Fig. 1a), indicating that this region might be involved in important protein–protein interactions.

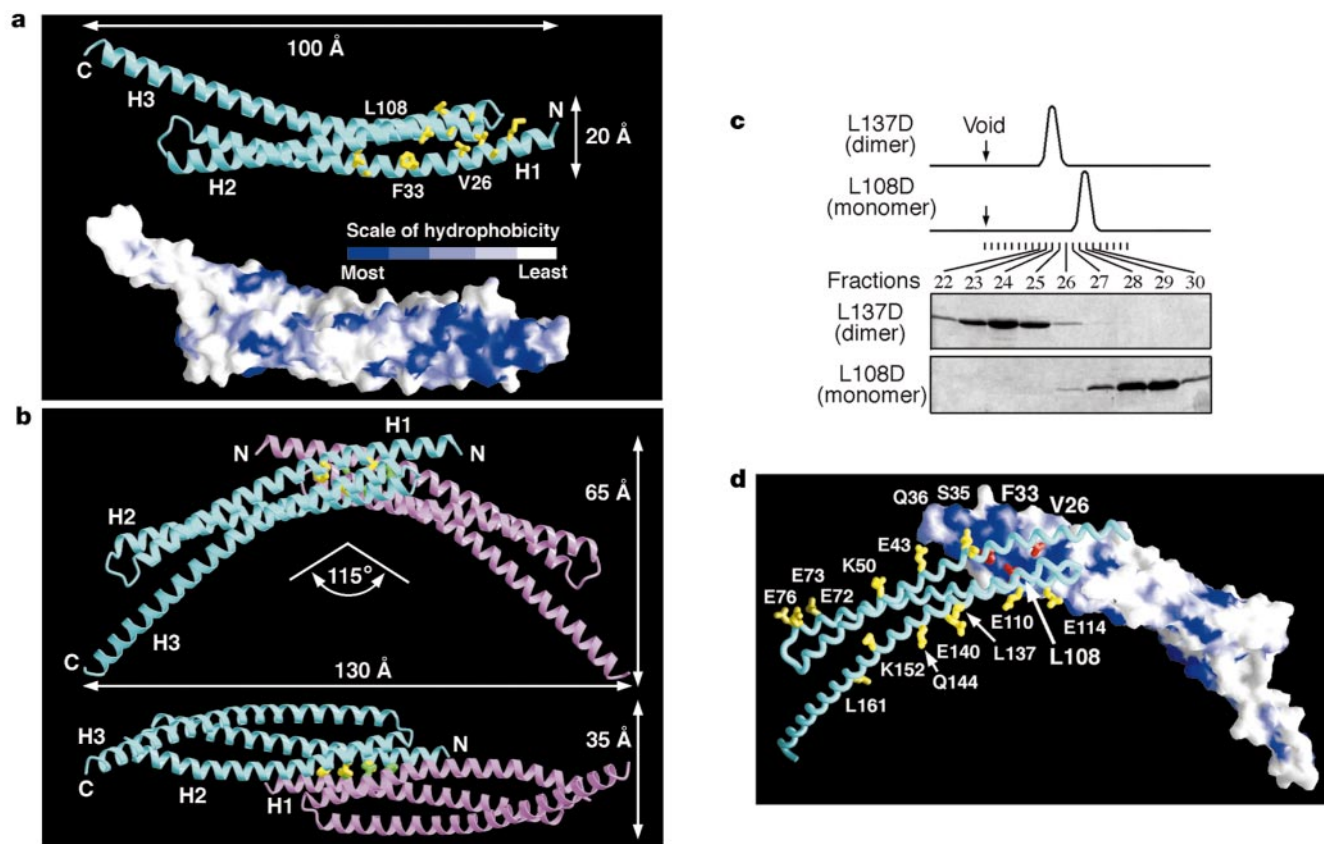


Figure 1 Schematic representation of Smac structure. **a**, The elongated structure of a Smac monomer. The hydrophobic surface patch is dark blue (bottom) and its constituent residues are yellow (top). **b**, The arch-shaped structure of a Smac dimer, with the two monomers coloured cyan and pink, respectively. The two views of the structure are related by a 90° rotation around a horizontal axis. Three critical residues at the dimeric interface, V26, F33 and L108, are coloured yellow and green for two Smac molecules. **c**, Representative chromatographs of Smac dimer (L137D) and monomer (L108D) mutants on gel filtration. We applied 0.5 mg of each of the 22 missense mutant proteins to

gel-filtration chromatography, and the protein-containing fractions were visualized by SDS–polyacrylamide gel electrophoresis. **d**, Mapping of residues targeted for missense mutation on Smac structure. Mutation of the three red residues, V26, F33 and L108, completely disrupted Smac dimers, whereas mutation of all other residues (yellow) had no detectable effect. The two Smac monomers are shown in the same orientation as in **b** except one is represented by hydrophobic surface. Figs 1 and 2 were prepared using MOLSCRIPT³⁶ and GRASP³⁷.

Overall structure of a Smac dimer

Smac was reported to behave as an oligomer in solution, with an apparent M_r of 100K (ref. 20). Because Smac exhibits the same oligomeric state in both physiological buffer and crystallization solution, the oligomeric interface of Smac is probably preserved in crystal packing. We therefore carefully examined the crystal packing of one protomer against its neighbours. Unexpectedly, we failed to identify any protomer–protomer interface that can be propagated to form a higher-order oligomer, indicating that Smac may be a dimer in solution. Our analysis identified a strong candidate dimer interface.

In the crystals, two protomers of Smac pack symmetrically across the previously identified hydrophobic surface patch formed by helices H1 and H2, resulting in the burial of 2,157 Å² surface area (Fig. 1b). Dimerization through this interface creates a further elongated and arch-shaped Smac molecule, with the N termini on top and the carboxy termini at the feet of the arch (Fig. 1b). The angle subtended by the two legs of the arch is around 115°. There is an extensive flat surface underneath the arch formed by the C-terminal third of helix H2 and the N-terminal third of helix H3. The overall dimensions of the Smac dimer are length 130 Å, height 65 Å and width 35 Å, as measured by the corresponding backbone Cα atoms. This structural arrangement would cause a Smac dimer to

exhibit a much greater apparent molecular mass in gel-filtration chromatography than the predicted theoretical weight.

To define the dimeric interface conclusively, we generated 22 missense mutations for specific surface residues in Smac, 15 of which targeted crystal packing contacts (Table 2). In particular, seven mutations were directed at residues at the observed dimeric interface (Table 2). All 22 mutant proteins were purified to homogeneity, and the oligomeric states of these proteins were individually analysed by gel-filtration chromatography (Fig. 1c, Table 2). Four point mutations (V26D, F33A, F33D and L108D), affecting three residues at the candidate dimeric interface, completely disrupted the oligomeric state of the wild-type protein (Fig. 1c, Table 2). The elution volume of these four mutants on gel filtration corresponds to an apparent M_r of about 33K, consistent with that expected for an elongated monomer (21K). In contrast, the other 18 mutations, including eight targeting other crystal packing contacts, had no detectable effect on the oligomeric state of Smac (Table 2). Because of the extensive and hydrophobic nature of the dimeric interface, mutation of a hydrophobic residue to alanine (V26A or L108A), or of a peripheral polar residue (Q36A), is not sufficient to abolish dimer formation (Fig. 1d, Table 2). These results confirm that Smac forms a symmetric dimer (Fig. 1b).

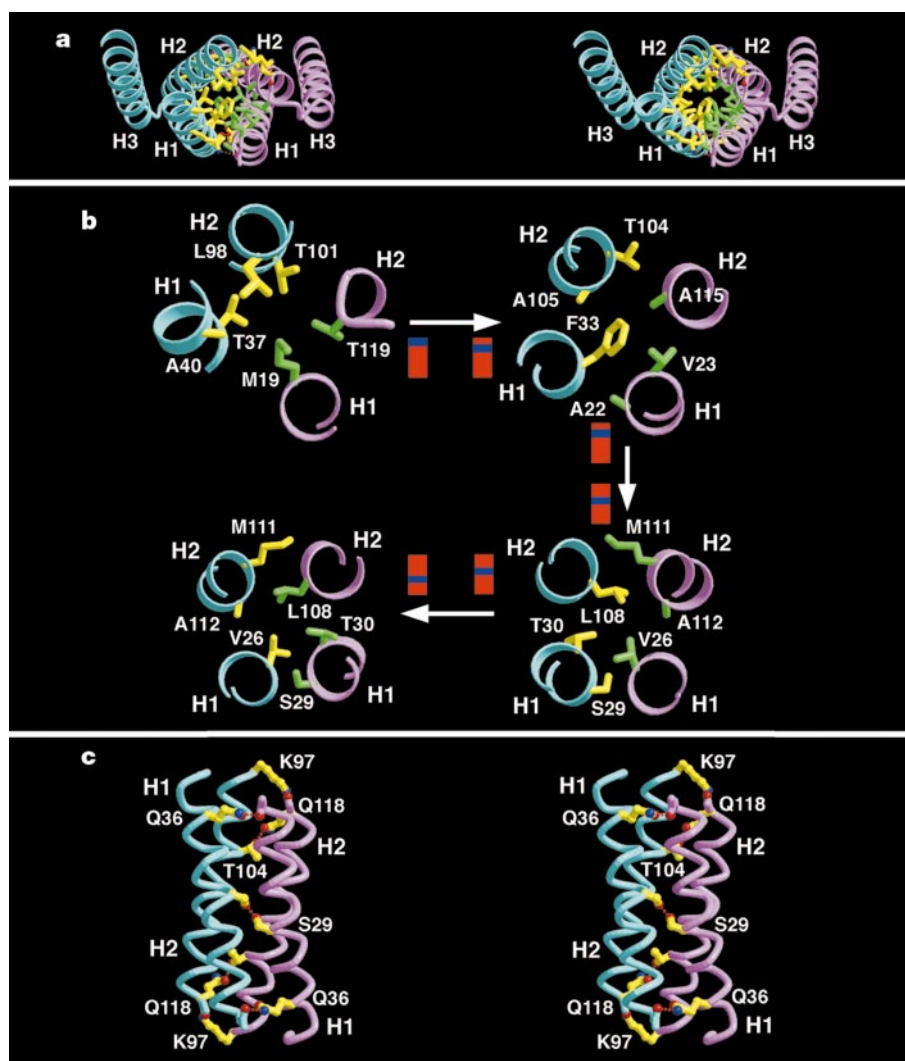


Figure 2 Specificity at the dimeric interface. **a**, Stereo diagram of the dimeric interface in Smac. The two Smac molecules are coloured cyan and pink. Interface residues from these two molecules are highlighted in yellow and green, respectively. Hydrogen bonds are represented by red dashed lines. **b**, Close-up views of the hydrophobic interactions at the

interface. The four panels show four consecutive planes of interface packing from the outside to the centre of the four-helix bundle. The orientation of helices and the colour scheme are as in **a**. **c**, Stereo representation of the seven hydrogen bonds, represented by red dashed lines, at the dimeric interface.

Specificity at the dimeric interface

Two molecules of Smac homodimerize through their N-terminal halves of helix H1 and C-terminal thirds of helix H2, forming an antiparallel four-helix bundle (Fig. 2a). The core interface is predominantly hydrophobic, with additional specificity provided by intermolecular hydrogen bonds at the periphery of the four-helix bundle (Fig. 2).

The hydrophobic packing at the dimeric interface is extensive, with 18 residues from one molecule interacting with those from the other molecule (Fig. 2a, b). The interface exhibits twofold symmetry, with each half consisting of three planes of interdigitating residues (Fig. 2b). Hydrophobic residues stack closely against each other both within and between adjacent planes. At the end of the four-helix bundle lies the first plane, where two Thr residues, T37 and T101, on molecule 1, pack against M19 and T119 on molecule 2 (Fig. 2b). Farther along the axis of the four-helix bundle is the second plane, where F33 on H1 extends into the centre (Fig. 2b). Five small residues, A105 and T104 on molecule 1 and A22, V23 and A115 on molecule 2, form a hydrophobic environment around F33. The third plane is located next to the centre of the four-helix bundle (Fig. 2b). Within this plane, L108 on molecule 1 and V26 on molecule 2 pack closely against each other at the centre. Four additional residues, T30 and S29 on helix H1 of molecule 1 and M111 and A112 on helix H2 of molecule 2, pack at the periphery of this plane, forming a network of van der Waals interactions.

Three residues, V26, F33 and L108, reside in the centre of the four-helix bundle and constitute the core of the hydrophobic interface (Fig. 2b). This structural arrangement predicts that replacement of these residues by charged residues is thermodynamically unfavourable and probably destabilizes the dimeric interface. Indeed, V26D, F33D and L108D all led to complete disruption of the dimeric interface (Fig. 1c, Table 2). Replacement of the bulky

F33 by a small Ala residue also abrogated dimer formation (Table 2).

Although hydrophobic contacts are important in the formation of a Smac dimer, hydrogen-bond interactions may further strengthen the interface and probably contribute to the specificity. There are seven intermolecular hydrogen bonds at the interface, all located at the periphery (Fig. 2c). At the end of the four-helix bundle, the side chain of K97 donates a hydrogen bond to the backbone carbonyl oxygen of Q118, whereas its side chain accepts a hydrogen bond from T104 (Fig. 2c). The side chain of Q36 makes a hydrogen bond to the backbone carbonyl group of L18 (Fig. 2c). At the centre of the helix bundle, S29 makes a symmetric contact to S29 in the other molecule (Fig. 2c).

Interaction with IAPs

Smac stimulates activation of procaspase-3 by relieving inhibition by IAPs^{20,21}. All members of the IAP family contain at least one BIR (baculoviral IAP repeat) motif, and many contain three⁹. Recent experiments indicate that different BIR domains may exhibit distinct functions. During Fas-induced apoptosis, the endogenous XIAP is cleaved into two fragments, each with distinct specificity for caspases²². In addition, the second BIR domain (BIR2) of XIAP appears to be a potent suppressor of apoptosis and a direct inhibitor for caspases, whereas neither BIR1 nor BIR3 exhibit similar activity²³.

To characterize Smac–IAP interaction further, we generated a series of XIAP fragments as fusion proteins with glutathione S-transferase (GST; Fig. 3a) and assayed their interaction with Smac using purified recombinant proteins (Table 2). As expected, wild-type Smac interacts stably with the second or third BIR domain of XIAP (Table 2). In contrast, Smac exhibited no detectable interaction with XIAP-BIR1 despite its strong sequence homology to the other two BIR domains (Table 2). To examine whether Smac can

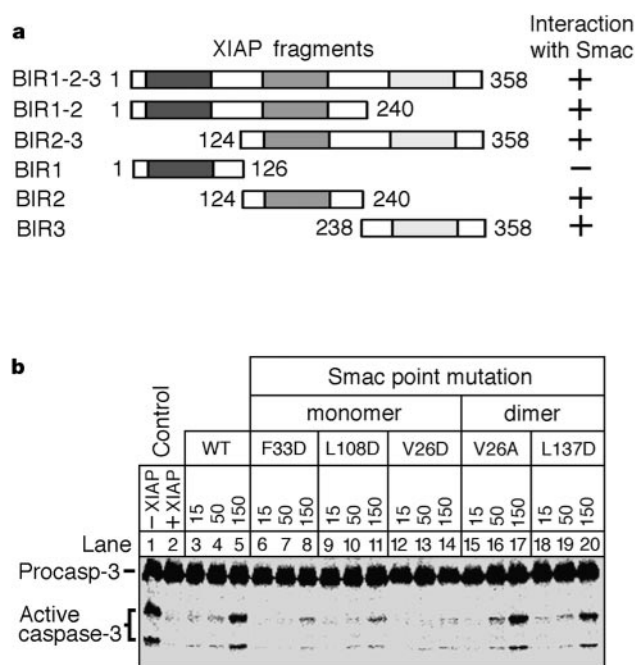
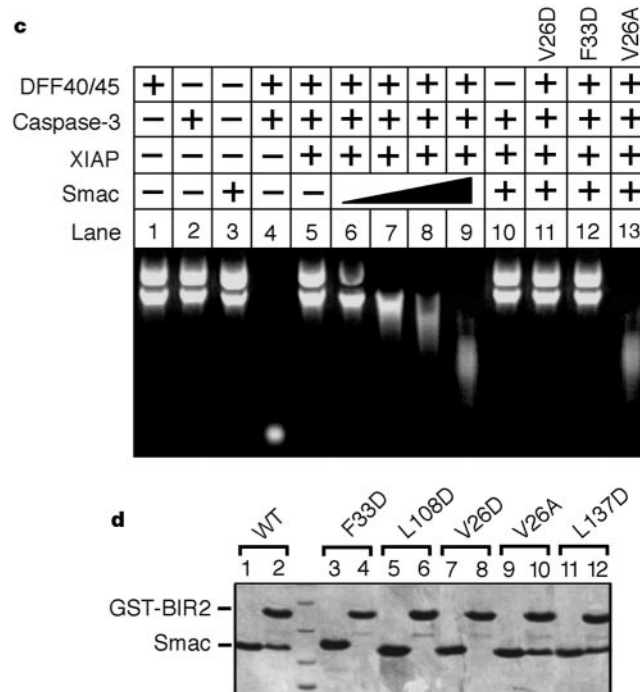


Figure 3 Functional significance of the dimeric interface in Smac. **a**, Schematic diagram of the XIAP fragments and results of their interaction with Smac. All fragments were expressed and purified as GST-fusion proteins. **b**, Missense mutations inactivating dimer formation exhibited severely diminished activation of procaspase-3. Lane 1, positive control in which procaspase-3 was converted to active caspase-3 in a totally reconstituted system containing purified recombinant Apaf-1, procaspase-9, cytochrome *c* and dATP. Lane 2, inhibition of procaspase-3 activation by XIAP. The same amount of XIAP was used for all other lanes. Numbers indicate Smac concentrations (nM). **c**, Smac missense



mutations that inactivate dimer formation fail to release XIAP inhibition of caspase-3 activity. The amounts of wild-type Smac were 3, 0.1, 0.3, 1, 3 and 3 μ g for lanes 3, 6, 7, 8, 9 and 10, respectively. We used 3 μ g of each of the mutant proteins for lanes 11–13. **d**, Missense mutations inactivating dimer formation disrupted interaction with the BIR2 domain of XIAP. The interaction was examined by GST-mediated pull-down assays. For each Smac mutant, the left and right lanes indicate the input protein and the final complex, respectively.

bind simultaneously to BIR2 and BIR3 of XIAP, we performed competition experiments. The results show that BIR2 and BIR3 exclude each other upon binding to Smac (data not shown).

In contrast to the wild-type dimeric Smac, the monomeric mutants V26D, F33D and L108D could not interact with the BIR2 domain of XIAP (Table 2). A fourth monomeric mutant, F33A, had severely diminished binding activity for XIAP-BIR2 (Table 2). All dimeric Smac mutants retained their ability to interact with XIAP-BIR2 (Table 2). Surprisingly, both monomeric and dimeric missense mutants can interact with XIAP-BIR3 (Table 2). In addition to its significance for the mechanism of caspase activation, this result also implies distinct specificity in the function of different BIR domains.

Dual effects of Smac

Smac can induce activation of procaspase-3 by eliminating the inhibitory effect of IAPs²⁰. To characterize this function *in vitro*, we reconstituted a procaspase-3 activation assay, using purified recombinant components (Fig. 3b). In the presence of Apaf-1, procaspase-9, cytochrome *c* and dATP, the radio-labelled procaspase-3 precursor was converted to the active form consisting of two subunits (Fig. 3b, lane 1). Addition of the recombinant XIAP protein (residues 1–356) completely inhibited the proteolytic processing of procaspase-3 (lane 2). The inhibitory effect of XIAP was eliminated by increasing amounts of wild-type Smac (lanes 3–5).

Apoptosis is carried out by the enzymatic activity of mature caspases. Although Smac induces the activation of procaspase-3, it is not clear whether it can also promote the catalytic activity of mature caspases. To investigate this possibility, we used a caspase-3 enzymatic assay, in which caspase-3 activity is monitored through its cleavage of DFF45 (ref. 24) or ICAD²⁵, the inhibitory subunit of the caspase-activated DNase. In apoptotic cells, degradation of DFF45 or ICAD releases inhibition of the DNase DFF40 or CAD, which subsequently cleaves chromosomes into nucleosomes^{25,26}. In the absence of caspase-3, the DFF complex remains inactive and the DNA substrate remains intact (Fig. 3c, lane 1). Incubation with active caspase-3 releases the DNase activity of DFF40, which

degrades the DNA substrate into short oligonucleotides (Fig. 3c, lane 4). Here the enzymatic activity of caspase-3 is measured as its ability to activate DFF40/45, which in turn cleaves DNA. Addition of XIAP-BIR2 to this reaction completely inhibits the enzymatic activity of caspase-3, as indicated by the absence of DNA cleavage (Fig. 3c, lane 5). Neither BIR1 nor BIR3 has any effect (data not shown), presumably because neither can bind caspase-3 (ref. 23). With the addition of increasing amounts of wild-type Smac, the inhibitory effect of XIAP on caspase-3 was relieved and the DNA was cleaved into progressively lower molecular mass species (Fig. 3c, lanes 6–9).

These observations show that Smac promotes the enzymatic activity of caspase-3 by physically removing the inhibitory effect of IAPs. Our results demonstrate that Smac has two roles: to induce the activation of procaspase-3, and to promote the enzymatic activity of mature caspase-3.

Functional significance of a Smac dimer

We first examined the role of the Smac dimer in the induction of procaspase-3 activation using the *in vitro* reconstituted assay (Fig. 3b). In contrast to wild-type Smac (Fig. 3b, lanes 3–5), the three monomeric Smac mutants, V26D, F33D and L108D, had markedly less capacity to promote procaspase-3 activation (Fig. 3b, lanes 6–14). Two dimeric Smac mutants, V26A and L137D, retained similar activity to wild-type Smac (Fig. 3b, lanes 15–20). These results indicate that Smac dimerization is essential to its function in activating procaspase-3.

Next we investigated the significance of a Smac dimer in promoting the enzymatic activity of mature caspase-3 (Fig. 3c). In contrast to wild-type Smac (Fig. 3c, lanes 6–9), the monomeric Smac mutants could not relieve the inhibition of caspase-3 activity by XIAP (Fig. 3c, lanes 11, 12). The activity of the dimeric mutant V26A was similar to that of the wild-type Smac (Fig. 3c, lane 13). Thus Smac promotes the enzymatic activity of caspase-3, and Smac dimerization is essential to this activity.

The functional dependence on dimerization of Smac can be explained by existing evidence and our interaction data (Fig. 3d).

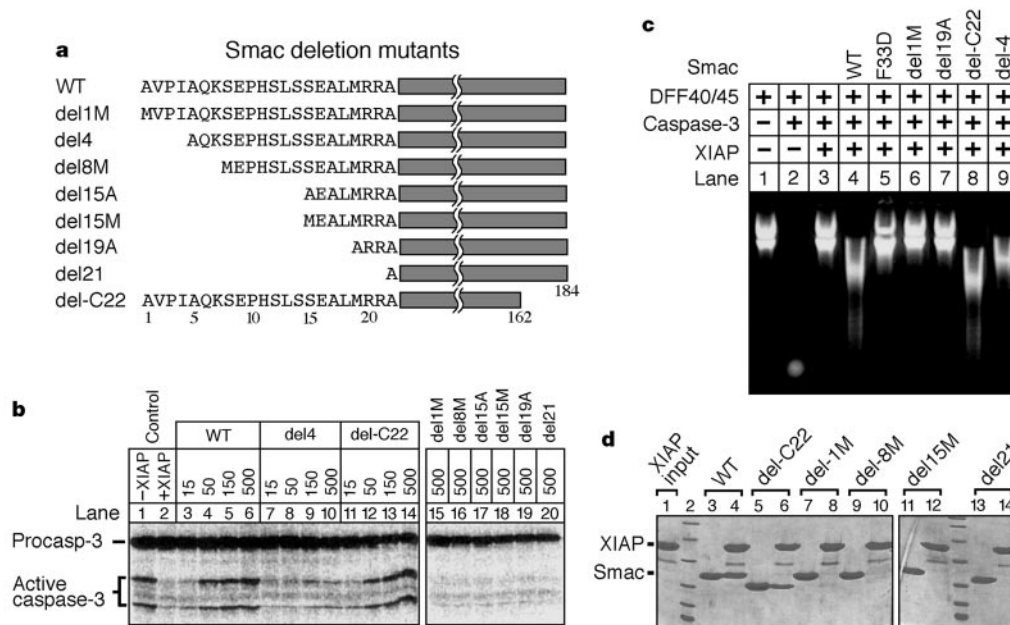


Figure 4 Functional significance of the N terminus of Smac. **a**, Smac N-terminal deletion mutants. The indicated sequences of N termini were confirmed by peptide sequencing. **b**, N-terminal deletion mutants exhibited severely diminished activation of procaspase-3. The concentrations of wild-type and mutant Smac are indicated in nM. **c**, N-terminal deletion mutants failed to remove the inhibition of caspase-3 activity by XIAP. The amount

of wild-type or mutant Smac protein used was 3 μ g (lanes 4–9). **d**, N-terminal deletion mutants cannot interact with the XIAP BIR2. The interaction was examined using GST-mediated pull-down assays. For each Smac mutant, the left and right lanes indicate the input protein and the final complex, respectively.

The second BIR domain of XIAP is uniquely potent in blocking the activation of procaspase-9 and in inhibiting the enzymatic activity of mature caspase-3 (ref. 23). However, monomeric Smac mutants failed to interact with XIAP-BIR2 (Fig. 3d, Table 2), thus making these mutants unable to induce procaspase-3 activation and promote the enzymatic activity of mature caspase-3 by relieving the inhibitory effect of XIAP.

Function of N-terminal residues in Smac

In the course of preparing recombinant proteins for biochemical assays, we discovered that Smac protein derived from an N-terminal GST fusion is completely inactive even after removal of GST by proteolysis. This result indicates that the N-terminal flexible sequences in Smac may be critical for its activity.

To investigate systematically the role of the N-terminal sequences, we generated a series of Smac mutants with N-terminal deletions (Fig. 4a). All mutant proteins were purified to homogeneity. The identities of these mutants were confirmed by mass spectroscopic analysis and N-terminal peptide sequencing. The first residue in endogenous wild-type Smac is Ala, owing to cleavage of the mitochondria targeting sequence²⁰. The initiation Met residue is completely removed in bacterial expression owing to the presence of an Ala as the penultimate residue²⁷ (Fig. 4a). We attempted to create a single-residue deletion by removing this Ala. However, the initiation Met is no longer removed owing to the presence of a Val as the penultimate residue²⁷. In essence, a single point mutation (Ala to Met) was created instead and this mutant was named del1M. The rest of the mutants were similarly created and named (Fig. 4a).

We assayed these N-terminal deletion mutants for their ability to promote caspase-3 activation (Fig. 4b). The missense mutation at the first residue, del1M, completely eliminated Smac activity (Fig. 4b, lane 15). All other deletion mutants, with the exception of del4, also lost their ability to promote procaspase-3 activation (Fig. 4b, lanes 16–20). Interestingly, the mutant in which the N-terminal four residues were removed (del4) was partially active (Fig. 4b, lanes 7–10). This mutation resulted in clean removal of the first four residues and thus partially preserved the identity of the N-

terminal sequence (Fig. 4a). Removal of the C-terminal 22 residues in Smac (del-C22) had no impact on its ability to promote caspase-3 activation (Fig. 4b, lanes 11–14).

We assayed these N-terminal deletion mutants for their ability to promote the enzymatic activity of caspase-3 (Fig. 4c). Wild-type Smac relieved the inhibitory effect of XIAP (Fig. 4c, lanes 4), but the Smac N-terminal deletion mutants were completely inactive in this assay (Fig. 4c, lanes 6, 7). In contrast, the activity of the C-terminal deletion mutant (del-C22) was indistinguishable from that of wild-type Smac (Fig. 4c, lane 8); del4 was also partially active (Fig. 4c, lane 9).

These results show that the N-terminal sequences are required for Smac function. None of the deletions affects the dimeric interface; and all deletion mutants examined behave as homodimers in solution (Table 2), ruling out the possibility that the loss of function for these deletions was due to disruption of the dimeric interface. Because Smac works by interacting with IAPs, we hypothesized that these N-terminal deletions may have removed its ability to interact with IAPs. None of the loss-of-function deletion mutants could interact with XIAP (Fig. 4d, Table 2). Consistent with its weak activity, the mutant del4 retained weak interaction with XIAP (Table 2).

N-terminal peptides activate procaspase-3

To examine whether the N-terminal peptides themselves might be able to promote activation of procaspase-3, several Smac peptides were chemically synthesized, purified, and assayed for their function (Fig. 5). The assay was identical to that shown in Fig. 3b, with the same set of control experiments (Fig. 5b, lanes 1–6). For comparison, we also included the result for a Smac monomer mutant (Fig. 5b, lanes 7–9). A peptide consisting of the first seven residues of Smac, Smac-7, could promote procaspase-3 activation at around 10 μM, and this activity reached its maximum at 300 μM (Fig. 5b, lanes 11–14). The same result was obtained for Smac-10 (Fig. 5b, lanes 16–19), Smac-11 and Smac-16 (data not shown). In contrast, Smac-7M, which differs from Smac-7 only in its first residue (Fig. 5a), had no detectable activity at 300 μM (Fig. 5b, lane 15). Another negative control peptide, Smac-7R, which corresponds to the reversed version of Smac-7, was also inactive for procaspase-3 activation (Fig. 5b, lane 10).

Table 2 Properties of Smac mutants

Smac mutant	Oligomeric state	Interaction with XIAP			
		BIR1 (1–126)	BIR2 (124–240)	BIR3 (238–358)	BIR1-2-3 (1–358)
WT	dimer	No	Yes	Yes	Yes
V26D*	monomer	No	No	Yes	Yes
V26A*	dimer	No	Yes	Yes	Yes
F33D*	monomer	No	No	Yes	Yes
F33A*	monomer	No	weak	Yes	Yes
Q36A*	dimer	No	Yes	Yes	Yes
L108D*	monomer	No	No	Yes	Yes
L108A*	dimer	No	Yes	Yes	Yes
S35A†	dimer	No	Yes	Yes	Yes
E43A†	dimer	No	Yes	Yes	Yes
L137D†	dimer	No	Yes	Yes	Yes
L137A†	dimer	No	Yes	Yes	Yes
L140D†	dimer	No	Yes	Yes	Yes
L140A†	dimer	No	Yes	Yes	Yes
Q144A†	dimer	No	Yes	Yes	Yes
L161A†	dimer	No	Yes	Yes	Yes
del1M	dimer	No	No	No	No
del4	dimer	No	weak	weak	weak
del8M	dimer	No	No	No	No
del15A	N/A	No	No	No	No
del15M	N/A	No	No	No	No
del19A	N/A	No	No	No	No
del21	N/A	No	No	No	No
del-C22	dimer	No	Yes	Yes	Yes

* These missense mutations target residues in the described dimeric interface (see text).
† These missense mutations target residues involved in other crystal packing contacts.
The following seven missense mutations targeting surface residues in Smac do not affect dimer formation: K50E, E72R, E73K, E76R, E110R, E114K and K152E.

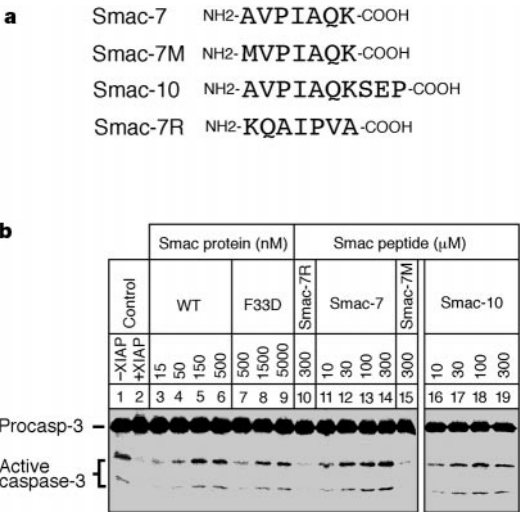


Figure 5 N-terminal peptides of Smac directly promote the activation of procaspase-3. **a**, The peptide sequences used in the assay. Smac-7 contains the N-terminal seven residues of Smac whereas Smac-7R represents the reversal of the sequence. **b**, Results of procaspase-3 activation assays. For comparison, the results of wild-type Smac protein and the F33D mutant are also shown (lanes 3–9).

Discussion

Despite high sequence similarity among the three BIR domains of XIAP, Smac interacts only with BIR2 and BIR3, not BIR1, indicating that the different DIR domains may have distinct functions. In addition, BIR2 and BIR3 exclude each other upon binding to Smac, indicating that they may recognize the same structural motif on Smac. N-terminal deletion mutants of Smac could not interact with either BIR domain (Table 2). Interestingly, the integrity of the dimeric interface is indispensable for interaction with BIR2 but not BIR3, indicating that the BIR2 domain might form a dimer in solution. In agreement with this proposal, BIR2 but not BIR3 can inhibit the catalytic activity of mature caspase-3 (ref. 23), which is a homodimer by itself²⁸. Further supporting this hypothesis, the BIR motifs of the IAP protein Op-IAP mediate self-oligomerization²⁹. Nevertheless, XIAP-BIR2 appeared to be a monomer in solution³⁰.

Interaction with IAPs depends on the integrity of the N terminus of Smac. This interaction appears to be highly specific, because any modification of the N-terminal residues in Smac compromised or abolished interaction. To our surprise, the replacement of the N-terminal residue Ala by Met in Smac completely abrogated interaction with IAPs, indicating that the N-terminal hydrophobic

residues in Smac may fit tightly into a surface groove on the BIR domain (Fig. 6a). In this case, the mutation of a small Ala residue by a bulky Met could abrogate binding through steric hindrance. This hypothesis is consistent with the observation that the Smac mutant del4, which lacks the first four residues, retains binding (Table 2). All binding studies between Smac and XIAP have been reproduced using c-IAP1 fragments with identical results (data not shown). Thus it is likely that Smac interacts with XIAP, c-IAP1 and c-IAP2 in a similar manner (Fig. 6a).

Although indispensable, the N-terminal sequences of Smac do not appear to be sufficient for stable interaction with IAPs. First, Smac monomeric mutants, such as F33D and V26D, maintain an intact N terminus but do not bind BIR2. Because the N terminus of one Smac protomer lies near the other protomer, the additional binding sites probably include regions of this other molecule (Fig. 6a). Second, although the Smac peptides were functional, they achieved a similar level of caspase-3 activation at higher concentrations compared with the Smac monomeric mutants (Fig. 5).

Previous studies indicated that Smac can promote the activation of procaspase-3 by eliminating the inhibitory effect of IAPs²⁰. We have shown that Smac promotes apoptosis through at least two mechanisms: inducing the proteolytic activation of procaspase-3 and promoting the enzymatic activity of mature caspase-3 (Fig. 6b). Both functions depend on the ability of Smac to interact physically with IAPs.

Smac functions as a dimer. Monomeric Smac exhibited 3–5-fold lower activation of procaspase-3 (Fig. 3b) but was largely inactive in promoting the enzymatic activity of mature caspase-3 (Fig. 3c). This is probably because monomeric Smac abrogated the interaction with the XIAP-BIR2 domain that is dominant in inhibiting caspase-3 activity²³. As expected, the Smac peptides can promote the activation of procaspase-3 only at concentrations higher than the monomeric mutants (Fig. 5b).

The *Drosophila* IAP protein, DIAP1, suppresses apoptosis by inhibiting caspases. Three *Drosophila* proteins, Reaper, Grim and Hid, induce apoptosis by eliminating this inhibitory effect through physical interactions^{11,12}. Thus Smac appears to be the mammalian functional homologue of Reaper, Grim and Hid. We have provided further evidence that the mechanisms by which Smac and Reaper/Grim/Hid activate apoptosis are conserved. Both Smac and Reaper/Grim/Hid contain an N-terminal fragment that is important for their function and for interactions with IAPs. The sequence homology among Reaper, Grim and Hid is restricted to their N-terminal 14 amino acids; deletion of these residues leads to loss of interaction with IAPs⁷. In addition, an N-terminal 37-residue peptide of Hid was sufficient to induce apoptosis and inhibit caspases in insect cells¹¹.

We also identified a 7-amino-acid peptide that can activate procaspase-3 activation. The high concentrations of this peptide required may be related to the amount of recombinant XIAP used in this experiment. Although untested, it is possible that even shorter peptides could possess similar activities. Thus, these peptides could be modified to test their ability to induce apoptosis in cancer cells that overexpress IAPs. □

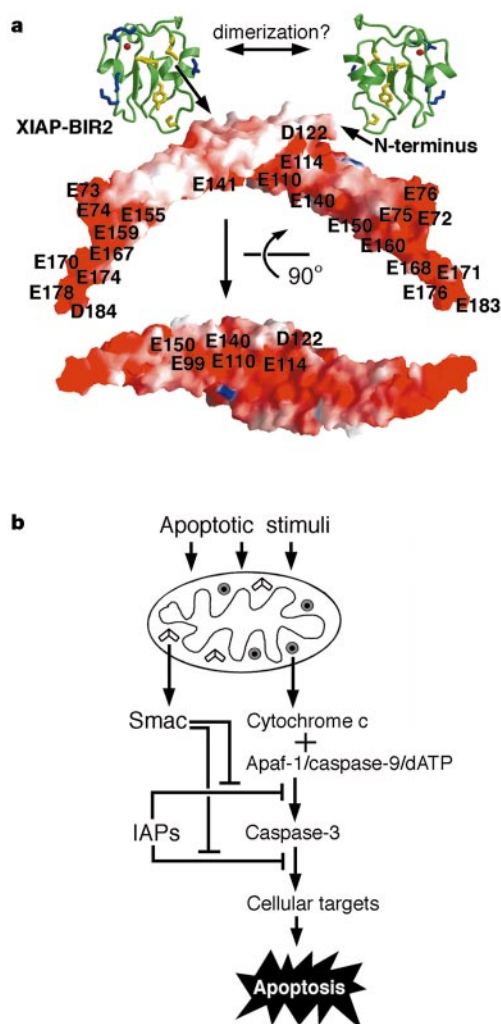


Figure 6 Model of Smac function. **a**, Model of interaction between Smac and IAPs. The surface of the Smac molecule is coloured by electrostatic potential, with negative and positive charges in red and blue, respectively. The BIR domain of IAPs, represented by XIAP-BIR2³⁰, is proposed to bind mainly the N-terminal sequences in Smac with additional contacts in the adjacent area. **b**, Model of Smac function. Smac not only induces the proteolytic activation of procaspase-3 but also promotes the catalytic activity of active caspase-3, both of which depend on its ability to bind IAPs.

Methods

Site-directed mutagenesis and protein preparation

We generated point mutations using a standard PCR-based cloning strategy, and verified the identities of individual clones through double-stranded plasmid sequencing. Both wild-type and mutant Smac were overexpressed in *Escherichia coli* strain BL21(DE3) as C-terminally 9-histidine-tagged proteins using a pET-15b vector (Novagen). Selenomethionyl Smac was expressed in *E. coli* B834(DE3) (Novagen) in M9 minimal medium supplemented with 50 mg l⁻¹ selenomethionine. The soluble fraction of the fusion protein in the *E. coli* lysate was purified over a Ni-NTA (Qiagen) column, and further fractionated by anion-exchange (Source-15Q, Pharmacia) and gel-filtration chromatography (Superdex-200, Pharmacia). Recombinant XIAP and c-IAP1 fragments were overexpressed as GST-fusion proteins using pGEX-2T (Pharmacia). GST-IAP in the *E. coli* lysate was

purified over a glutathione sepharose column, and further purified by anion-exchange chromatography (Source-15Q).

DFF40/45 complexes were co-expressed in *Escherichia coli* strain BL21(DE3) and purified to homogeneity through anion-exchange (SP-sepharose and Source-15S) and gel-filtration chromatography (Superdex-200). Active caspase-3 was expressed as a C-terminally 6-histidine-tagged protein using a pET-21b vector (Novagen) and purified as described above.

In vitro interaction assay

Interactions between Smac and IAPs were examined by GST-mediated pull-down assays. About 0.4 mg of a recombinant IAP fragment was bound to 200 µl of glutathione resin as a GST-fusion protein and incubated with 0.5 mg of wild-type or mutant Smac at room temperature. After extensive washing with an assay buffer containing 25 mM Tris, pH 8.0, 150 mM NaCl and 2 mM dithiothreitol (DTT), the complex was eluted with 5 mM reduced glutathione and visualized by SDS-PAGE with Coomassie staining.

Assay for activation of procaspase-3

Procaspase-3 was translated and purified as described³¹. An aliquot of the *in vitro* translated procaspase-3 (1 µl) was incubated with relevant proteins in the presence of 10 µM dATP at 30 °C for 1 h in a final volume of 20 µl in buffer A (20 mM HEPES-KOH, pH 7.5, 10 mM KCl, 2.5 mM MgCl₂, 1 mM DTT and 0.1 mM PMSF). Each reaction contained 38 nM recombinant Apaf-1, 12 nM recombinant procaspase-9 and 500 nM purified horse heart cytochrome c. The mouse XIAP fragment (residues 1–356) was used at 40 nM concentration to inhibit the reaction. At the end of incubation, 7 µl of 4×SDS sample buffer was added to each reaction, which was analysed by SDS-PAGE and phosphorimaging. The XIAP fragment used in this assay is derived from mouse. XIAP fragments used for all other assays are of human origin.

Assay for enzymatic activity of mature caspase-3

The reaction was carried out at 37 °C for 20 min in a final volume of 20 µl in a buffer containing 25 mM Tris, pH 8.0, 150 mM NaCl, 5 mM MgCl₂ and 100 µg ml⁻¹ BSA. We used 0.25 µg of ICAD/CAD complex, 0.1 µg of active caspase-3 and 1 µg of GST-XIAP-BIR2 (residues 124–240) where indicated. In each reaction, 2 µg of pEMBL plasmid was used as a substrate for the activated CAD. The reactions were subjected to 2% agarose gel electrophoresis and stained with ethidium bromide.

Crystallization and data collection

Crystals were grown at 4 °C by the hanging-drop vapour-diffusion method by mixing Smac protein (15 mg ml⁻¹) with an equal volume of reservoir solution containing 20% 1,4-dioxane (v/v), 200 mM ammonium sulphate and 10 mM DTT. The crystals grew over 3–4 weeks. The crystals are in the primitive hexagonal space group P6₃22, with unit cell dimensions $a = b = 108.51$ Å, $c = 70.38$ Å, and contain one Smac molecule in each asymmetric unit. Diffraction data were collected using an R-Axis-IV imaging plate detector mounted on a Rigaku 200HB generator. Derivatives were obtained by soaking crystals in appropriate heavy atom solutions under the condition of a cryoprotectant buffer containing 20% 1,4-dioxane (v/v), 200 mM ammonium sulphate and 20% glycerol (v/v). The concentration and soaking time for EMP, HgAc, K₃UO₂F₆, PhNH₂HgAc, MeHgAc and HgThimerosal were 0.5 mM/18 h, 0.5 mM/12 h, 1 mM/20 h, 2 mM/35 h, 2 mM/24 h and 0.5 mM/10 h, respectively.

Structure determination

The heavy atom positions of the EMP, HgAc and selenium derivative were determined using SOLVE³² and further refined using MLPHARE³³. The heavy atom sites of all other derivatives were identified using the difference Fourier method. Initial MIR phases calculated with the program MLPHARE³³ had a mean figure of merit of 0.586 to 3.0 Å resolution, and were improved with solvent flattening and histogram matching using program DM³³. A model was built into MIR electron density maps with program O³⁴ and refined by simulated annealing using CNS³⁵. The final refined atomic model contains Smac residues 11–182 and 47 water molecules. The N-terminal ten residues and the C-terminal two residues in Smac have no electron density in the maps, and we presume that these regions are disordered in the crystals.

Received 4 July; accepted 26 July 2000.

1. Steller, H. Mechanisms and genes of cellular suicide. *Science* **267**, 1445–1449 (1995).
2. Jacobson, M. D., Weil, M. & Raff, M. C. Programmed cell death in animal development. *Cell* **88**, 347–354 (1997).
3. Hengartner, M. O. Programmed cell death in invertebrates. *Curr. Opin. Genet. Dev.* **6**, 34–38 (1996).
4. Horvitz, H. R. Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. *Cancer Res.* **59**, 1701–1706 (1999).
5. Thompson, C. B. Apoptosis in the pathogenesis and treatment of disease. *Science* **267**, 1456–1462 (1995).

6. Green, D. R. & Martin, S. J. The killer and the executioner: how apoptosis controls malignancy. *Curr. Opin. Immunol.* **7**, 694–703 (1995).
7. Thornberry, N. A. & Lazebnik, Y. Caspases: Enemies within. *Science* **281**, 1312–1316 (1998).
8. Chinnaiyan, A. M. & Dixit, V. M. The cell-death machine. *Curr. Biol.* **6**, 555–562 (1996).
9. Deveraux, Q. L. & Reed, J. C. IAP family proteins—suppressors of apoptosis. *Genes Dev.* **13**, 239–252 (1999).
10. Miller, L. K. An exegesis of IAPs: salvation and surprises from BIR motifs. *Trends Cell Biol.* **9**, 323–328 (1999).
11. Wang, S., Hawkins, C., Yoo, S., Muller, H.-A. & Hay, B. The *Drosophila* caspase inhibitor DIAP1 is essential for cell survival and is negatively regulated by HID. *Cell* **98**, 453–463 (1999).
12. Goyal, L., McCall, K., Agapite, J., Hartwig, E. & Steller, H. Induction of apoptosis by *Drosophila* reaper, hid and grim through inhibition of IAP function. *EMBO J.* **19**, 589–597 (2000).
13. Zou, H., Henzel, W. J., Liu, X., Lutschg, A. & Wang, X. Apaf-1, a human protein homologous to *C. elegans* CED-4, participates in cytochrome c-dependent activation of caspase-3. *Cell* **90**, 405–413 (1997).
14. Li, P. et al. Cytochrome c and dATP-dependent formation of Apaf-1/Caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479–489 (1997).
15. Srinivasula, S. M., Ahmad, M., Fernandes-Alnemri, T. & Alnemri, E. S. Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization. *Mol. Cell* **1**, 949–957 (1998).
16. Yang, X., Chang, H. Y. & Baltimore, D. Essential role of CED-4 oligomerization in CED-3 activation and apoptosis. *Science* **281**, 1355–1357 (1998).
17. Hu, Y., Ding, L., Spencer, D. M. & Nunez, G. WD-40 repeat region regulates Apaf-1 self-association and procaspase-9 activation. *J. Biol. Chem.* **273**, 33489–33494 (1998).
18. Zou, H., Li, Y., Liu, X. & Wang, X. An APAF-1-cytochrome c multimeric complex is a functional aptosome that activates procaspase-9. *J. Biol. Chem.* **274**, 11549–11556 (1999).
19. Saleh, A., Srinivasula, S. M., Acharya, S., Fishel, R. & Alnemri, E. S. Cytochrome c and dATP-mediated oligomerization of Apaf-1 is a prerequisite for procaspase-9 activation. *J. Biol. Chem.* **274**, 17941–17945 (1999).
20. Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
21. Verhagen, A. et al. Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing inhibitor of apoptosis (IAP) proteins. *Cell* **102**, 43–53 (2000).
22. Deveraux, Q. L. et al. Cleavage of human inhibitor of apoptosis protein XIAP results in fragments with distinct specificities for caspases. *EMBO J.* **18**, 5242–5251 (1999).
23. Takahashi, R. et al. A single BIR domain of XIAP sufficient for inhibiting caspases. *J. Biol. Chem.* **273**, 7787–7790 (1998).
24. Liu, X., Zou, H., Slaughter, C. & Wang, X. DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* **89**, 175–184 (1997).
25. Enari, M. et al. A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43–50 (1998).
26. Liu, X. et al. The 40-kDa subunit of DNA fragmentation factor induces DNA fragmentation and chromatin condensation during apoptosis. *Proc. Natl Acad. Sci. USA* **95**, 8461–8466 (1998).
27. Hirel, P.-H., Schmitter, J.-M., Dessen, P., Fayat, G. & Blanquet, S. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl Acad. Sci. USA* **86**, 8247–8251 (1989).
28. Rotonda, J. et al. The three-dimensional structure of apopain/CPP32, a key mediator of apoptosis. *Nature Struct. Biol.* **3**, 619–625 (1996).
29. Hozak, R. R., Manji, G. A. & Friesen, P. D. The BIR motifs mediate dominant interference and oligomerization of inhibitor of apoptosis Op-IAP. *Mol. Cell Biol.* **20**, 1877–1885 (2000).
30. Sun, C. et al. NMR structure and mutagenesis of the inhibitor-of-apoptosis protein XIAP. *Nature* **401**, 818–822 (1999).
31. Liu, X., Kim, C. N., Yang, J., Jemmerson, R. & Wang, X. Induction of apoptosis program in cell-free extracts: Requirement for dATP and cytochrome c. *Cell* **86**, 147–157 (1996).
32. Terwilliger, T. C. & Berendzen, J. Correlated phasing of multiple isomorphous replacement data. *Acta Crystallogr. D* **52**, 749–757 (1996).
33. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
34. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
35. Brunger, A. T. et al. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
36. Klauis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
37. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Acknowledgements

We thank D. Vaux for providing XIAP cDNA; E. Alnemri for providing caspase-3 expression vector; F. Hughson for critically reading the manuscript; and N. Hunt for secretarial assistance. This work was supported by start-up funds from Princeton University (to Y.S.) and Howard Hughes Medical Institute (to X.W.). Y.S. is a Searle Scholar and a Rita Allen Scholar.

Correspondence and requests for materials should be addressed to Y.S. (e-mail: yshi@molbio.princeton.edu). The atomic coordinates have been deposited with the Protein Data Bank with the accession number 1FEW.

Cooper instability of composite fermions

Vito W. Scarola, Kwon Park & J. K. Jain

Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

When confined to two dimensions and exposed to a strong magnetic field, electrons screen the Coulomb interaction in a topological fashion; they capture an even number of quantum vortices and transform into particles called 'composite fermions' (refs 1–3). The fractional quantum Hall effect⁴ occurs in such a system when the ratio (or 'filling factor', ν) of the number of electrons and the degeneracy of their spin-split energy states (the Landau levels) takes on particular values. The Landau level filling $\nu = 1/2$ corresponds to a metallic state in which the composite fermions form a gapless Fermi sea^{5–8}. But for $\nu = 5/2$, a fractional quantum Hall effect is observed instead^{9,10}; this unexpected result is the subject of considerable debate and controversy¹¹. Here we investigate the difference between these states by considering the theoretical problem of two composite fermions on top of a fully polarized Fermi sea of composite fermions. We find that they undergo Cooper pairing to form a p -wave bound state at $\nu = 5/2$, but not at $\nu = 1/2$. In effect, the repulsive Coulomb interaction between electrons is overscreened in the $\nu = 5/2$ state by the formation of composite fermions, resulting in a weak, attractive interaction.

The most important property of composite fermions is that they do not experience the external magnetic field B but rather a drastically reduced magnetic field $B^* = B - 2p\rho\phi_0$. Here ρ is the two-dimensional density of fermions, $2p$ is the number of vortices carried by the composite fermion, often intuitively envisioned as $2p$ flux quanta, and $\phi_0 = h/e$ is a flux quantum. In effect, each electron absorbs $2p$ flux quanta of the external field to turn into a composite fermion. Electrons confined in two dimensions have unusual properties in a magnetic field; considerable progress towards understanding these properties has been made by modelling the composite fermion (CF) system as a non-interacting gas of composite fermions^{1,2}. In particular, the fractional quantum Hall effect (FQHE)⁴ is a manifestation of the integral quantum Hall effect of composite fermions³, and the metallic, non-FQHE state at the half-filled lowest Landau level is a Fermi sea of composite fermions^{5–8}.

The Landau level (LL) filling $\nu = 5/2 = 2 + 1/2$ corresponds to a half-filled second LL. Here, both spin states of the lowest LL are completely occupied, contributing 2 to the filling factor. The fully

occupied LL is treated as inert in this work, and the electrons in the partially filled second LL are assumed to be fully spin-polarized. These are valid approximations in sufficiently high magnetic fields. In complete analogy to the half-filled lowest LL, the model of non-interacting composite fermions would predict a Fermi sea of composite fermions at $\nu = 5/2$ as well. However, experiments reveal a FQHE state here^{9,10}. In fact, $5/2$ is the only even-denominator fraction to be observed in a single-layer system, and its physical origin has been a subject of debate and controversy¹¹.

To understand the fundamental difference between $\nu = 5/2$ and $\nu = 1/2$, it is necessary to go beyond the model of non-interacting composite fermions. Our theoretical investigations of the inter-CF interaction employ the Jain wavefunctions for composite fermions^{1–3}. These wavefunctions not only give an accurate quantitative account of the inter-CF interaction, but even capture the subtle, interaction-driven Wigner and Bloch instabilities of the CF liquid at small filling factors^{12,13}; such instabilities are analogous to those believed to occur for the ordinary electron gas (jellium) at low densities¹⁴.

Central to this work is the following question, analogous to the Cooper problem for ordinary superconductivity: if we begin by assuming a Fermi sea of composite fermions both at $\nu = 1/2$ and $\nu = 5/2$ and add two composite fermions at the Fermi surface, will they form a bound state? We find that the CF-Fermi sea is unstable to pairing of composite fermions at $\nu = 5/2$ but not at $\nu = 1/2$, as shown schematically in Fig. 1. We stress that, in contrast to the Bardeen–Cooper–Schrieffer (BCS) theory, we do not assume any attractive interaction, phonon-mediated or otherwise; the only interaction in the problem is the repulsive Coulomb interaction between electrons. However, the Coulomb interaction translates into a weak attractive interaction between composite fermions at $\nu = 5/2$.

We work in the spherical geometry^{15,16}, in which N electrons are considered to move on the surface of a sphere under the presence of a radial magnetic field produced by a magnetic monopole of

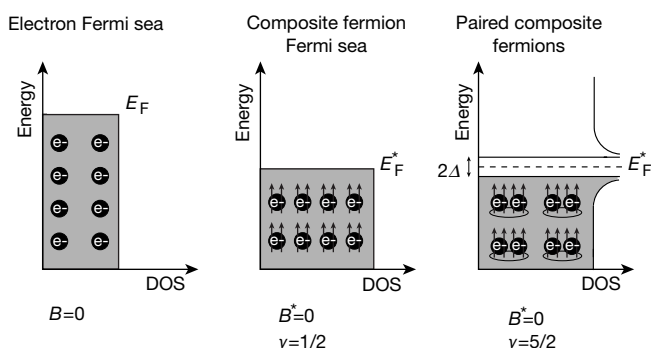


Figure 1 Density of states (DOS) for electrons and composite fermions at zero effective magnetic flux. Left panel, electrons at $B = 0$. Centre panel, composite fermions at $\nu = 1/2$. Right panel, pairing of composite fermions at $\nu = 5/2$. The composite fermions are shown as electrons carrying two flux quanta.

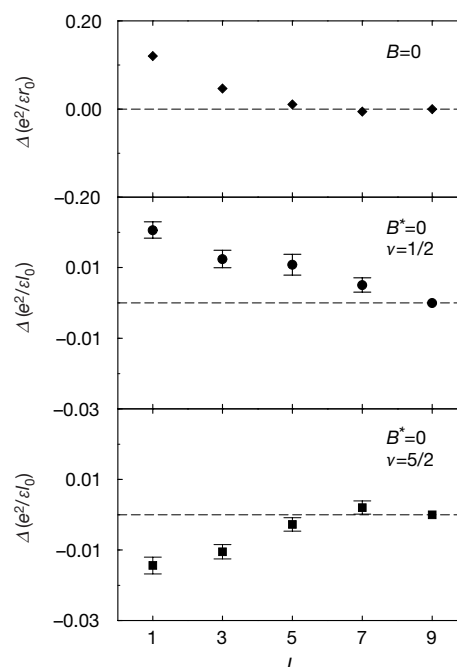


Figure 2 The interaction energy of pairs of electrons and pairs of composite fermions at zero effective magnetic flux as a function of angular momentum L . The results are shown for a system with $N = 27$ particles. Top panel, electrons; middle panel, composite fermions at $\nu = 1/2$; bottom panel, composite fermions of $\nu = 5/2$. The quantity $l_0 = \sqrt{h/eB}$ is the magnetic length, $r_0 = (\pi\rho)^{-1/2}$ is the average interparticle separation, and ϵ is the dielectric constant of the background material ($\epsilon = 12.8$ for GaAs).

strength Q at the centre. The flux through the surface of the sphere is $2Q\phi_0$, where $2Q$ is an integer, according to Dirac's quantization condition. The composite fermion theory maps the problem of interacting electrons at Q to that of composite fermions at $Q^* = Q - N + 1$ (assuming here and below composite fermions of vorticity $2p = 2$). The Jain wavefunctions for interacting electrons at Q are given by $\Psi_Q = P_{LL} \Phi_1^2 \Phi_{Q^*}$, where Φ_{Q^*} are wavefunctions of non-interacting electrons at Q^* , Φ_1 is the wavefunction of the fully occupied lowest Landau level at $Q_1 = (N - 1)/2$, and P_{LL} is the lowest LL projection operator. Due to the rotational symmetry, the total orbital angular momentum L is a good quantum number, preserved in going from Φ_{Q^*} to Ψ_Q .

We are interested in composite fermions in a vanishing effective magnetic field, that is, when $Q^* = 0$, which is obtained at $Q = N - 1$. Here, for $N = n^2$, the ground state Ψ_Q (Φ_{Q^*}) has uniform density ($L = 0$), as it contains n filled shells of composite fermions (electrons). We approach the CF Fermi sea as the $N \rightarrow \infty$ limit of the filled-shell states. The systems with a CF pair occur at $Q^* = 0$ for $N = n^2 + 2$, corresponding to the situation when two CF particles are added to the $(n + 1)$ st shell. The individual angular momenta of the additional particles are $l = n$, implying that there is one multiplet at each total angular momentum $L = 1, 3, \dots, L_{\max} = 2n - 1$ with a degeneracy of $2L + 1$. The wavefunction of a CF pair at $B^* = 0$ for a given L is $\Psi_L^{\text{CF-pair}} = P_{LL} \Phi_1^2 \Phi_L^{\text{el-pair}}$, where $\Phi_L^{\text{el-pair}}$ is the wavefunction of an electron pair at $B = 0$. We will also consider a pair of CF holes, corresponding to two holes in the n th shell at $Q^* = 0$. $\Psi_L^{\text{CF-pair}}$ contains no adjustable parameter.

For $\Phi_L^{\text{el-pair}}$, the Coulomb energy of the electron pair, which is proportional to the average inverse distance between the two electrons in the otherwise empty shell, decreases with L , in accordance with Hund's rule of atomic physics¹⁷. This implies that the smallest L ($L = 1$ for fully polarized particles) corresponds to the smallest distance between two electrons of a pair. By analogy, the size of the CF pair in $\Psi_L^{\text{CF-pair}}$ also increases with L . We have confirmed this by monitoring the influence of an additional short-range interaction to the pair energy.

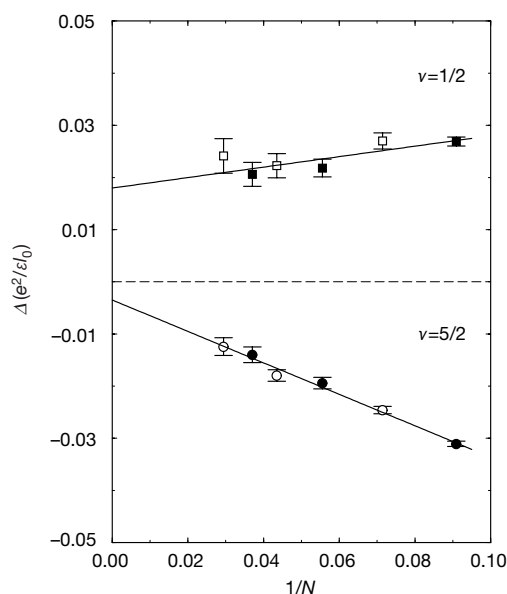


Figure 3 The binding energy of the composite fermion (CF) pair in the $L = 1$ channel at $\nu = 1/2$ and $5/2$ as a function of $1/N$. The binding energy is defined as the energy of the pair at $L = 1$ relative to its energy at $L_{\max}$. The filled (empty) symbols are for CF-particles (CF-holes) on top of the CF-Fermi surface. As expected from particle-hole symmetry, satisfied by the wavefunctions considered here to an excellent approximation, the binding energies for the CF-particles and CF-holes fall on the same line.

As it stands, $\Psi_L^{\text{CF-pair}}$ is written in the lowest Landau level, that is, for $\nu = 1/2$. In order to treat $\nu = 5/2$, we use the method of Park *et al.*¹⁸ to map the problem of the Coulomb interaction in the second ($s = 1$) Landau level into that of an effective interaction $V^{\text{eff}}(r)$ in the lowest ($s = 0$) Landau level. $V^{\text{eff}}(r)$ is chosen so that the two interactions have the same Haldane pseudopotentials¹⁵, that is, $V_{1,m} = V_{0,m}^{\text{eff}}$. The parameters $V_{s,m}$ are the interaction energies of two particles in s th LL in relative angular momentum m state (in the planar geometry) and completely specify the interparticle interaction in the s th Landau level.

The energy of $\Psi_L^{\text{CF-pair}}$, $E(L)$, is evaluated in a Monte Carlo approach¹⁹. Figure 2 shows the energy of the pair as a function of L for electrons at $B = 0$, and for composite fermions at $\nu = 1/2$ and $5/2$. The interaction energy between the two added electrons decreases with increasing L at $B = 0$, as expected. Similar behaviour is found for composite fermions at $\nu = 1/2$. However, the opposite behaviour is seen at $\nu = 5/2$, indicating an attractive interaction between composite fermions at $\nu = 5/2$. The largest binding energy is obtained for the $L = 1$ channel. This qualitative difference between $\nu = 1/2$ and $\nu = 5/2$ is the principal result of this work.

The estimation of the thermodynamic limit of the binding energy from our finite-size study can be difficult. When the pair interaction energy is not small compared to the inter-shell spacing, many shells would participate in the pair wavefunction, and our approximation of restricting the pair to the lowest unoccupied shell would break down. A determination of the appropriate 'parent' wavefunction $\Phi^{\text{el-pair}}$ is obviously quite complicated in this regime, and consequently, so is obtaining $\Psi^{\text{CF-pair}}$. However, past studies²⁰ have shown that distinct excitations in Φ with the same quantum numbers may produce the same excitation Ψ , because the Hilbert space at $\nu = 1/2$ is greatly restricted compared to that at zero magnetic field. Therefore, as a first step, we proceed without any explicit consideration of shell mixing. We have studied up to $n = 6$ filled shells; both CF-particle pairs and CF-hole pairs are considered for up to $n = 5$ and only the latter for $n = 6$. Figure 3 shows that despite a substantial decrease with N , the binding energy $\Delta[L = 1] = E[L = 1] - E[L_{\max}]$ at $\nu = 5/2$ still extrapolates to a non-zero negative value of -0.0035 (± 0.0013) $e^2/\epsilon l_0$ as $N \rightarrow \infty$, where $l_0 = \sqrt{\hbar/eB}$ and ϵ is the dielectric constant of the background material. After taking account of the transverse width of the electron wavefunction^{21,22}, the binding energy in the $L = 1$ channel is estimated to be -0.0025 (± 0.0015) $e^2/\epsilon l_0$ for densities in the range $(0.5-3.0) \times 10^{11} \text{ cm}^{-2}$. Following the weak-coupling BCS theory, it is natural to identify $2|\Delta|$ with the energy gap of the FQHE state at $\nu = 5/2$. For the parameters of the experiment of Pan *et al.*¹⁰, our calculated $2|\Delta|$ for $L = 1$ is 0.5 (± 0.3) K, which turns out to be on the same order of magnitude as the measured gap of about 0.1 K. While this is encouraging, we note that the effect of shell mixing is an important issue, and must be considered to ascertain the quantitative reliability of our preliminary estimate.

The appearance of pairing may seem surprising in a model with strong repulsive interaction. However, the Coulomb repulsion is overcome through the formation of composite fermions, which screens out the Coulomb interaction quite effectively, to the extent that a total neglect of the interaction between composite fermions is sufficient for many purposes. Furthermore, the screening takes place in a topologically rigid manner, independent of the interaction strength or the Landau level index, through the binding of precisely two vortices to each electron. Therefore, it is plausible that sometimes an overscreening of the Coulomb interaction occurs, producing an effectively attractive interaction between composite fermions. The reason that there is an attraction at $\nu = 5/2$ but not at $\nu = 1/2$ is that the matrix elements for the Coulomb interaction in the second Landau level are weaker than in the lowest LL because of the greater spread of the electron wavefunction in the former, especially at short distances. For example, $V_{1,1}/V_{0,1} = 0.93$. This slight softening of the inter-electron repulsive interaction in the

second LL is sufficient to make the inter-CF interaction weakly negative.

There has been earlier work on pairing of composite fermions. Greter, Wen and Wilczek²³ argued for *p*-wave pairing of composite fermions at $\nu = 1/2$ and $5/2$ within a Chern–Simons formulation of composite fermions. The Chern–Simons method, however, is quantitatively unreliable in this application because of its inadequacy in describing the energetics or the short-distance behaviour. Even within this approach, Bonesteel²⁴ has noted that a pair-breaking term not considered in ref. 23 may potentially alter its conclusion. Greiter *et al.* further suggested that the paired CF state may be described in terms of a Pfaffian wavefunction written by Moore and Read²⁵. Recent exact diagonalization^{26,27} and variational¹⁸ studies have provided support for the validity of a Pfaffian-like wavefunction at $\nu = 5/2$.

The pairing of composite fermions at $\nu = 5/2$ has a topological origin, and occurs in spite of strong repulsive interaction between electrons. The repulsion is circumvented because the objects that pair up are not electrons but composite fermions. We speculate that a fundamental reorganization of the state, for example, the creation of new quasiparticles, must happen in any system in order for pairing to ensue, starting from repulsive interactions alone. This is indeed the case in several theoretical models of high-temperature superconductivity, where the pairing is believed also to be caused by repulsive interactions. □

Received 6 March 2000; accepted 15 June 2000.

- Heinonen, O. (ed.) *Composite Fermions* (World Scientific, New York, 1998).
- Das Sarma, S. & Pinczuk, A. (eds) *Perspectives in Quantum Hall Effects* (Wiley, New York, 1997).
- Jain, J. K. Composite-fermion approach for the fractional quantum Hall effect. *Phys. Rev. Lett.* **63**, 199–202 (1989).
- Tsui, D. C., Stormer, H. L. & Gossard, A. C. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
- Halperin, B. I., Lee, P. A. & Read, N. Theory of the half-filled Landau level. *Phys. Rev. B* **47**, 7312–7343 (1993).
- Willett, R. L. *et al.* Experimental demonstration of a Fermi surface at one-half filling of the lowest Landau level. *Phys. Rev. Lett.* **71**, 3846–3849 (1993).
- Goldman, V. J. *et al.* Detection of composite fermions by magnetic focusing. *Phys. Rev. Lett.* **72**, 2065–2068 (1994).
- Kang, W. *et al.* How real are composite fermions? *Phys. Rev. Lett.* **71**, 3850–3853 (1993).
- Willett, R. L. *et al.* Observation of an even-denominator quantum number in the fractional quantum Hall effect. *Phys. Rev. Lett.* **59**, 1779–1782 (1987).
- Pan, W. *et al.* Exact quantization of the even-denominator fractional quantum Hall state at $\nu = 5/2$ Landau level filling factor. *Phys. Rev. Lett.* **83**, 3530–3533 (1999).
- Eisenstein, J. P. in *Perspectives in Quantum Hall Effects* (eds Das Sarma, S. & Pinczuk, A.) 37–70 (Wiley, New York, 1997).
- Kamilla, R. K. & Jain, J. K. Excitonic instability and termination of fractional quantum Hall effect. *Phys. Rev. B* **55**, R13417–R13420 (1997).
- Park, K. & Jain, J. K. Spontaneous magnetization of composite fermions. *Phys. Rev. Lett.* **83**, 5543–5546 (1999).
- Tanatar, B. & Ceperley, D. M. Ground state of the two-dimensional electron gas. *Phys. Rev. B* **39**, 5005–5016 (1989).
- Haldane, F. D. M. Fractional quantization of the Hall effect: a hierarchy of incompressible quantum fluid states. *Phys. Rev. Lett.* **51**, 605–608 (1983).
- Wu, T. T. & Yang, C. N. Dirac monopole without string: Monopole harmonics. *Nucl. Phys. B* **107**, 365–380 (1976).
- Wojs, A. & Quinn, J. J. Hund's rule for monopole harmonics, or why the composite fermion picture works. *Solid State Commun.* **110**, 45–49 (1999).
- Park, K., Melik-Alaverdian, V., Bonesteel, N. E. & Jain, J. K. Possibility of *p*-wave pairing of composite fermions at $\nu = 1/2$. *Phys. Rev. B* **58**, R10167–R10170 (1998).
- Jain, J. K. & Kamilla, R. K. Composite fermions in the Hilbert space of the lowest electronic Landau level. *Int. J. Mod. Phys. B* **11**, 2621–2660 (1997).
- Wu, X. G. & Jain, J. K. Excitation spectrum and collective modes of composite fermions. *Phys. Rev. B* **51**, 1752–1761 (1995).
- Fang, F. F. & Howard, W. E. Negative field-effect mobility on (100) Si surfaces. *Phys. Rev. Lett.* **16**, 797–800 (1966).
- Zhang, F. C. & Das Sarma, S. Excitation gap in the fractional quantum Hall effect: Finite layer thickness corrections. *Phys. Rev. B* **33**, 2903–2906 (1986).
- Greiter, M., Wen, X. G. & Wilczek, F. Paired Hall states. *Nucl. Phys. B* **374**, 567–614 (1992).
- Bonesteel, N. E. Singular pair breaking in the composite Fermi liquid description of the half-filled Landau level. *Phys. Rev. Lett.* **82**, 984–987 (1999).
- Moore, G. & Read, N. Nonabelions in the fractional quantum Hall effect. *Nucl. Phys. B* **360**, 362–396 (1991).
- Morf, R. Transition from quantum Hall to compressible states in the second Landau level: new light on the $\nu = 5/2$ enigma. *Phys. Rev. Lett.* **80**, 1505–1508 (1998).
- Rezaei, E. H. & Haldane, F. D. M. Transition from paired quantum Hall to compressible states at the half filling of the lowest two Landau levels. Preprint cond-mat/9906137 at (<http://xxx.lanl.gov>) (1999).

Acknowledgements

This work was supported in part by the National Science Foundation. We thank the Numerically Intensive Computing Group led by V. Agarwala, J. Holmes and J. Nucciarone, at the Penn State University CAC, for assistance and computing time with the LION-X cluster.

Correspondence and requests for materials should be addressed to J.K.J. (e-mail: jain@phys.psu.edu).

Nitride semiconductors free of electrostatic fields for efficient white light-emitting diodes

P. Waltereit, O. Brandt, A. Trampert, H. T. Grahn, J. Menniger, M. Ramsteiner, M. Reiche & K. H. Ploog

Paul-Drude-Institut für Festkörperelektronik, Hausvogteiplatz 5-7, D-10117 Berlin, Germany

Compact solid-state lamps based on light-emitting diodes (LEDs)^{1,2} are of current technological interest as an alternative to conventional light bulbs. The brightest LEDs available so far emit red light and exhibit higher luminous efficiency than fluorescent lamps. If this luminous efficiency could be transferred to white LEDs, power consumption would be dramatically reduced, with great economic and ecological consequences. But the luminous efficiency of existing white LEDs is still very low, owing to the presence of electrostatic fields within the active layers³. These fields are generated by the spontaneous and piezoelectric polarization along the [0001] axis of hexagonal group-III nitrides—the commonly used materials for light generation^{4–6}. Unfortunately, as this crystallographic orientation corresponds to the natural growth direction of these materials deposited on currently available substrates⁷. Here we demonstrate that the epitaxial growth of GaN/(Al,Ga)N on tetragonal LiAlO₂ in a non-polar direction allows the fabrication of structures free of electrostatic fields, resulting in an improved quantum efficiency. We expect that this approach will pave the way towards highly efficient white LEDs.

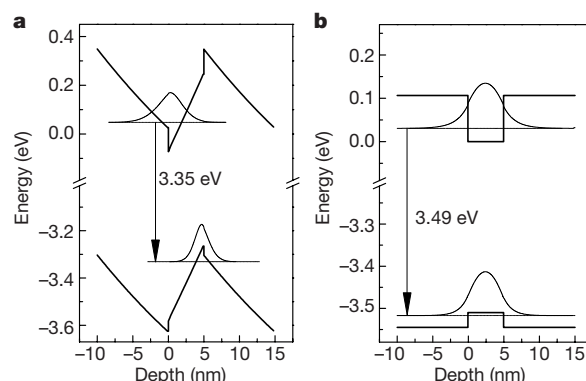


Figure 1 Calculated band profiles in (5 nm GaN)/(10 nm Al_{0.1}Ga_{0.9}N) quantum wells. These profiles were obtained by self-consistent effective mass Schrödinger–Poisson calculations. The transition energies given take into account both strain and Coulomb interaction. **a**, The very large electrostatic fields in the [0001] orientation (polarization charges were taken from ref. 4) result in a quantum confined Stark effect and poor electron–hole overlap. **b**, The [1100] orientation is free of electrostatic fields, thus true flat-band conditions are established.

Present white LEDs consist of a blue LED pumping a yellow phosphor¹. These LEDs only approach the luminous efficiency of incandescent lamps ($12\text{--}20\text{ lm W}^{-1}$). Furthermore, the emission of these LEDs is not perceived as true white by the eye, because the inevitable scattering of the pump light results in the impression of a bluish or yellowish colour⁸. An alternative and conceptually superior approach consists of an ultraviolet (UV) LED pumping a white phosphor. Here, the colour impression can be reproducibly adjusted by choosing the appropriate phosphor. Currently, however, the power efficiency of UV LEDs ($1\text{--}2\%$; ref. 9) is much less than that of blue LEDs ($10\text{--}12\%$).

Here we discuss the consequences of the very large electrostatic fields that exist within thin layers such as quantum wells. In Fig. 1a, we show the band profile of a GaN/(Al,Ga)N multiple quantum well structure, together with the ground-state wavefunctions for electrons and holes, calculated self-consistently within effective mass theory. The poor electron–hole overlap results in long radiative lifetimes¹⁰ and thus in low internal quantum efficiencies^{3,11}, given that competing non-radiative channels are always present at elevated temperature. The addition of indium to the quantum well, as in blue LEDs, increases the quantum efficiency because carriers are localized at the compositional fluctuations of the resulting inhomogeneous alloy⁹. This benefit, however, comes at the expense of an

unreproducible redshift in emission, which results in the deviation from truly white emission described above.

We now consider how these electrostatic fields might be removed. Doping, which might at first glance be considered to screen the fields, is no solution because the magnitude of the fields requires carrier concentrations in the high 10^{19} cm^{-3} range to achieve flat-band conditions^{6,12,13}. A more elegant alternative is to use the unpolarized ‘zinc blende’ phase of nitride semiconductors, which is not piezoelectric when grown along [001]. However, even after intense efforts¹⁴, the lack of suitable substrates and the inherent thermodynamic metastability of this phase has prohibited the achievement of device-quality material. This leaves us with the ‘wurtzite’ phase, which carries a spontaneous polarization even in the absence of strain.

A material is polarized at equilibrium if it has a singular polar axis: this is the case with the wurtzite structure, with its [0001] axis. The zinc blende structure, in contrast, cannot be polarized as it has four symmetry-equivalent polar axes, whose contributions cancel each other. The very uniqueness of the singular [0001] polar axis of the wurtzite phase implies that any direction orthogonal to it, such as $[1\bar{1}00]$ (ref. 15), will not carry spontaneous polarization components. At the same time these directions will carry no piezoelectric polarization if shear stresses in the growth plane are absent¹⁶, as is the case in our present study (see below). The band profile of our GaN/(Al,Ga)N multiple quantum well structure, grown along such a direction, would thus be unaffected by electrostatic fields. Figure 1b shows this flat-band case, which exhibits a higher transition energy and an electron–hole overlap of essentially unity.

There have been some failed attempts to grow GaN along directions perpendicular to the [0001] axis⁷. Due to the huge mismatch ($+49\%$) between GaN and the commonly used substrate, sapphire, the observed growth direction was the same as that found for the non-epitaxial deposition of GaN on glass: [0001].

We have succeeded in growing GaN($1\bar{1}00$) on tetragonal (γ) $\text{LiAlO}_2(100)$ (ref. 17) as originally postulated by Hellman *et al.*¹⁸. His results documented the persistence of GaN in growing only along the [0001] direction, even in the presence of another, favourable direction¹⁸. Indeed, in contrast to a conventional substrate such as 6H-SiC(0001), which offers hexagons as nucleation sites (Fig. 2a) for the hexagonal basal plane (‘C-plane’) of GaN with a lattice mismatch of 3.4% , the surface of $\gamma\text{-LiAlO}_2(100)$ (Fig. 2b) exhibits nucleation sites for the rectangular prism plane (‘M-plane’) of GaN with an anisotropic mismatch of only $1\text{--}2\%$. The [0001] axis then lies in the growth plane, fulfilling the requirement for the absence of electric polarization components along the growth axis.

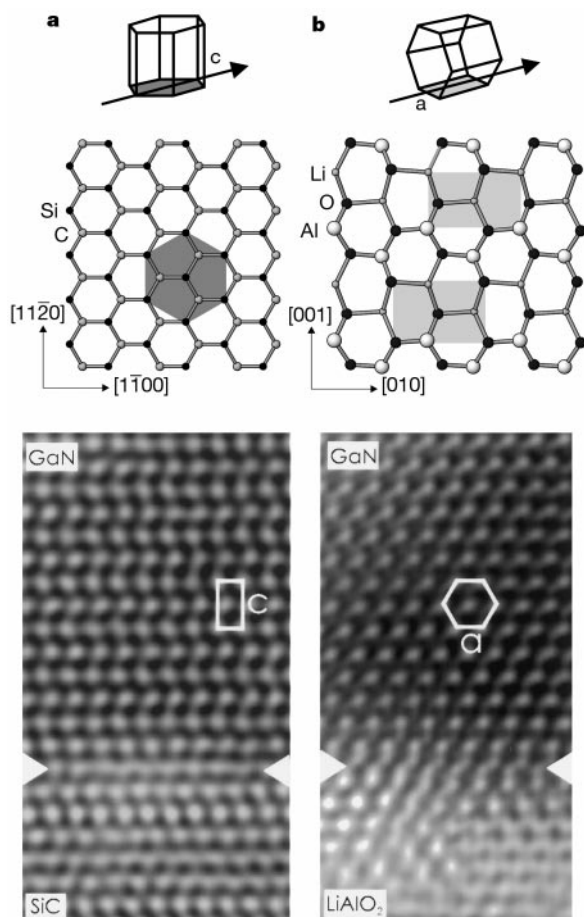


Figure 2 Structural characterization of hexagonal GaN. This figure shows schematic unit cells (top), nucleation sites in a ball-and-stick model (centre) and cross-sectional high-resolution transmission electron microscopy (HREM) images (bottom) for the growth of hexagonal GaN with lattice constants c and a . **a**, C-plane GaN(0001) on the Si-face of SiC(0001); **b**, M-plane GaN($1\bar{1}00$) and $\gamma\text{-LiAlO}_2(100)$ either on cations (upper rectangle) or anions (lower rectangle). Arrows (top) indicate the projected directions in HREM. We note the 2H stacking sequence of GaN(0001) grown on 6H-SiC(0001) and the hexagonal symmetry of GaN($1\bar{1}00$) grown on $\gamma\text{-LiAlO}_2(100)$ in HREM.

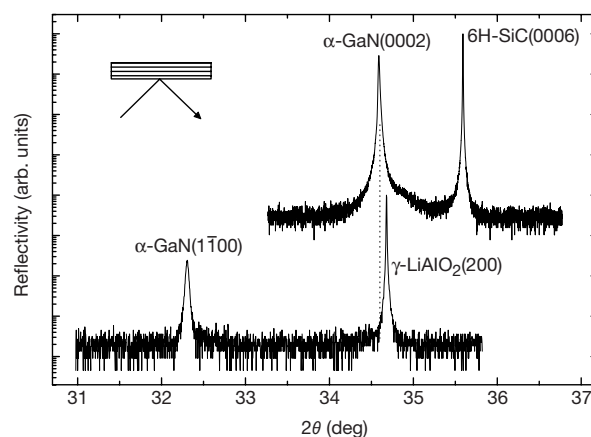


Figure 3 Symmetric triple-crystal $\omega\text{--}2\theta$ X-ray diffraction scans of a GaN($1\bar{1}00$) film grown on $\gamma\text{-LiAlO}_2(100)$ and a GaN(0001) film grown on 6H-SiC(0001). We note the absence of a GaN(0002) signal for the sample grown on $\gamma\text{-LiAlO}_2$.

We first study the orientation and phase purity of GaN layers grown on these two substrates by plasma-assisted molecular-beam epitaxy. The growth conditions for deposition on γ -LiAlO₂(100) are the same as those used for 6H-SiC(0001) (ref. 19) except for nucleation, which was performed at considerably lower substrate temperature (P.W., O.B., M.R. and K.H.P., unpublished results). Additionally, smooth and clean γ -LiAlO₂(100) substrates were found to be crucial to avoid the formation of GaN(0001).

The different orientation of layers grown on these substrates is directly seen by high-resolution transmission electron microscopy. Figure 2a shows the GaN(0001)/6H-SiC(0001) interface. The 6H stacking sequence of SiC is seen to abruptly change into the 2H stacking sequence of GaN(0001) along the growth direction. The GaN[0001] axis is thus perpendicular to the interface. In contrast, Fig. 2b shows the GaN(1100)/LiAlO₂(100) interface. The epilayer exhibits a hexagonal symmetry as the GaN[0001] axis is parallel to the interface. We used high-resolution X-ray diffraction scans to examine the phase composition of the two samples on a macroscopic scale (Fig. 3). For the layer grown on 6H-SiC(0001), we observe, as expected, a single reflection corresponding to C-plane GaN(0001), while the layer grown on γ -LiAlO₂(100) diffracts only at the angular position of M-plane GaN(1100). The complete absence of the C-plane GaN(0002) reflection means that this layer is single-phase GaN(1100). This result has been confirmed by Raman scattering (P.W., O.B., M.R. and K.H.P., unpublished results).

An important prerequisite for the growth of quantum wells is a smooth morphology of the layers. We found by atomic force microscopy that the surface morphology of GaN(1100) is comparable in smoothness to that of GaN(0001) (ref. 19). Thus encouraged, we grew identical 15-period (5 nm GaN)/(10 nm Al_{0.1}Ga_{0.9}N) multiple quantum well structures on 6H-SiC(0001) and γ -LiAlO₂(100) employing GaN buffer layers. The same structural parameters, verified by high-resolution X-ray diffraction, were used for the calculations in Fig. 1.

Figure 4 shows cathodoluminescence spectra of these samples taken at 5 K. The transition energy for the C-plane wells is 3.36 eV, whereas the M-plane wells emit at 3.48 eV. Considering the different strain states caused by the different substrates²⁰, we find that the C-plane wells emit 90 meV below, and the M-plane wells 10 meV above the GaN bandgap, respectively, in very satisfactory agreement with the calculated transition energies in Fig. 1. In particular, while the transition energy of the M-plane wells results solely from quantum confinement, that of the C-plane wells suggests the presence of strong electrostatic fields. The several well-defined peaks of the

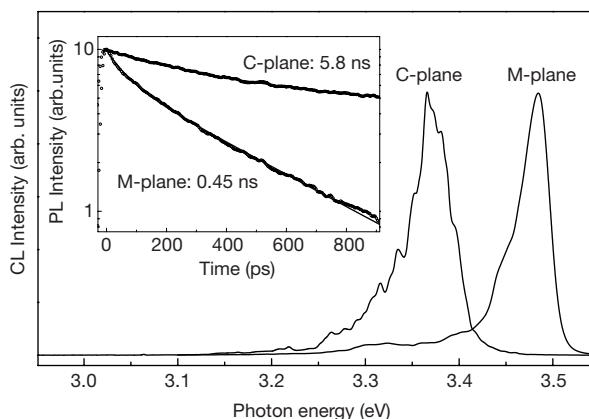


Figure 4 The effect of electric polarization on the emission characteristics of GaN/(Al,Ga)N multiple quantum wells (MQWs). The pronounced redshift in emission as well as the much prolonged decay times (inset) of the C-plane MQWs with respect to the field-free M-plane MQWs are both consequences of the internal electrostatic fields. CL, cathodoluminescence; PL, photoluminescence.

C-plane wells cannot be explained by thickness interference-fringes, taking into account the sample thickness of about 1.5 μ m. Instead, we propose that these peaks originate from excitons localized at steps and compositional fluctuations. Due to the electrostatic fields present in C-plane wells, localization here leads to much larger energy shifts than in the M-plane wells.

The effect of the electrostatic fields on the electron-hole overlap is evident from time-resolved photoluminescence measurements at 5 K (inset of Fig. 4). The striking difference between the decay times for M-plane (450 ps) and C-plane (≥ 6 ns) wells is a direct manifestation of the presence of electrostatic fields in the latter wells. At this low temperature, we can safely assume that the measured decay time corresponds to the value for the radiative lifetime, which, in turn, is inversely proportional to the overlap integral²¹. Depending on the (poorly known) material parameters, we arrive at 10–100 times longer radiative lifetimes for the C-plane wells, which is in fair agreement with the experiment. At room temperature, this long radiative lifetime results in a low quantum efficiency, as also reported by Deguchi *et al.*³ using commercial LEDs produced by Nichia.

The above results show how the internal quantum efficiency of GaN/(Al,Ga)N multiple quantum wells can be improved by using an unusual crystal orientation. We can, however, also improve the external quantum efficiency of the structure. In contrast to sapphire or SiC, γ -LiAlO₂ can be selectively etched with respect to GaN, and the epitaxial structure can thus be bonded onto an integrated reflector²², which enables the production of on-chip LEDs. In a second step, the GaN surface may be textured by dry chemical etching²³, to reduce total internal reflection at the GaN/air interface. We are currently pursuing this concept, and aim to obtain LED arrays with high internal and external quantum efficiencies. □

Received 25 February; accepted 4 July 2000.

- Ponce, F. A. & Bour, D. P. Nitride-based semiconductors for blue and green light-emitting devices. *Nature* **386**, 351–359 (1997).
- Morkoç, H. & Mohammad, S. N. High-luminosity blue and blue-green gallium nitride light-emitting diodes. *Science* **267**, 51–55 (1995).
- Deguchi, T. *et al.* Quantum-confined Stark effect in an AlGaIn/GaN/AlGaIn single quantum well structure. *Jpn J. Appl. Phys.* **38**, L914–L916 (1999).
- Bernardini, F., Fiorentini, V. & Vanderbilt, D. Spontaneous polarization and piezoelectric constants of III-V nitrides. *Phys. Rev. B* **56**, R10024–R10027 (1997).
- Bernardini, F. & Fiorentini, V. Macroscopic polarization and band offsets at nitride heterojunctions. *Phys. Rev. B* **57**, R9427–R9430 (1998).
- Florentini, V., Bernardini, F., Della Sala, F., Di Carlo, A. & Lugli, P. Effects of macroscopic polarization in III-V nitride multiple quantum wells. *Phys. Rev. B* **60**, 8849–8858 (1999).
- Iwata, K., Asahi, H., Asami, K., Kuroiwa, R. & Gonda, S. Gas source molecular beam epitaxy growth of GaN on C-, A-, R- and M-plane sapphire and silica glass substrates. *Jpn J. Appl. Phys.* **36**, L661–L664 (1997).
- Bogner, G. *et al.* White LED production at Osram. *Compound Semicond.* **5**, 28–31 (1999).
- Mukai, T., Morita, D. & Nakamura, S. High-power UV InGaIn/AlGaIn double-heterostructure LEDs. *J. Cryst. Growth* **189/190**, 778–781 (1998).
- Langer, R. *et al.* Giant electric fields in unstrained GaN single quantum wells. *Appl. Phys. Lett.* **74**, 3827–3829 (1999).
- Nishida, T. & Kobayashi, N. NTT reports 346 nm emission from AlGaIn LED. *Compound Semicond.* **5**, 12 (1999).
- Thamm, A. *et al.* Optical properties of heavily doped GaN/(Al,Ga)N multiple quantum wells grown on 6H-SiC(0001) by reactive molecular-beam epitaxy. *Phys. Rev. B* **61**, 16025–16028 (2000).
- Jiang, H. *et al.* Photoluminescence and photoluminescence excitation spectra of In_{0.2}Ga_{0.8}N-GaN quantum wells: comparison between experimental and theoretical studies. *IEEE J. Quant. Electron.* **35**, 1483–1490 (1999).
- Brandt, O. In *Group III Nitride Semiconductor Compounds: Physics and Applications* (ed. Gil, B.) 417–459 (Oxford Univ. Press, 1998).
- Northrup, J. E. & Neugebauer, J. Theory of GaN(1100) and (1120) surfaces. *Phys. Rev. B* **53**, R10477–R10480 (1996).
- Ludwig, W. *Festkörperphysik* (Akademische Verlagsgesellschaft, Wiesbaden, 1978).
- Marezio, M. The crystal structure and anomalous dispersion of γ -LiAlO₂. *Acta Crystallogr.* **19**, 396–400 (1965).
- Hellman, E. S., Lilienthal-Weber, Z. & Buchanan, D. N. F. Epitaxial growth and orientation of GaN on (100) γ -LiAlO₂. *MRS Internet J. of Nitride Research* **2**, 30 (<http://nsr.mij.mrs.org/2/>) (1997).
- Brandt, O. *et al.* Critical issues for the growth of high quality (Al,Ga)N/GaN and GaN/(In,Ga)N heterostructures on SiC(0001) by molecular beam epitaxy. *Appl. Phys. Lett.* **75**, 4019–4021 (1999).
- Waltereit, P. *et al.* Influence of AlN nucleation layers on growth mode and strain relief of GaN grown on 6H-SiC(0001). *Appl. Phys. Lett.* **74**, 3660–3662 (1999).
- Della Sala, F. *et al.* Free-carrier screening of polarization fields in wurtzite GaN/InGaIn laser structures. *Appl. Phys. Lett.* **74**, 2002–2004 (1999).

22. Schnitzer, I., Yablonovitch, E., Caneau, C., Gmitter, T. J. & Scherer, A. 30% external quantum efficiency from surface textured, thin-film light-emitting diodes. *Appl. Phys. Lett.* **63**, 2174–2176 (1993).
23. Fricke, J., Yang, B., Brandt, O. & Ploog, K. Patterning of cubic and hexagonal GaN by Cl_2/N_2 -based reactive ion etching. *Appl. Phys. Lett.* **74**, 3471–3473 (1999).

Acknowledgements

We thank O. Mayrock and H.-J. Wünsche for help in the calculation of single-particle wavefunctions and exciton binding energies. This work was supported in part by the Volkswagen-Stiftung.

Correspondence and requests for materials should be addressed to P.W. (e-mail: walter@pdi-berlin.de).

Liquid crystal phase transitions in suspensions of polydisperse plate-like particles

Felix M. van der Kooij, Katerina Kassapidou & Henk N. W. Lekkerkerker

Van't Hoff Laboratory for Physical and Colloid Chemistry, Debye Institute, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Colloidal suspensions that form periodic self-assembling structures on sub-micrometre scales are of potential technological interest; for example, three-dimensional arrangements of spheres in colloidal crystals¹ might serve as photonic materials², intended to manipulate light. Colloidal particles with non-spherical shapes (such as rods and plates) are of particular interest because of their ability to form liquid crystals. Nematic liquid crystals possess orientational order; smectic and columnar liquid crystals additionally exhibit positional order (in one or two dimensions respectively). However, such positional ordering^{3,4} may be inhibited in polydisperse colloidal suspensions. Here we describe a suspension of plate-like colloids that shows isotropic, nematic and columnar phases on increasing the particle concentration. We find that the columnar two-dimensional crystal persists for a polydispersity of up to 25%, with a cross-over to smectic-like ordering at very high particle concentrations. Our results imply that liquid crystalline order in synthetic mesoscopic materials may be easier to achieve than previously thought.

Suspensions of (almost) monodisperse colloidal particles are used as mesoscopic models of the phase behaviour of atomic and molecular systems. In this respect, the fluid–crystal transition in suspensions of spheres¹ and the liquid crystal phase transitions in suspensions of non-spherical particles^{5,6} have attracted considerable

attention. However, colloidal particles are hardly ever truly monodisperse but often exhibit a size distribution of finite width. The consequences of this inherent polydispersity are of substantial fundamental and industrial interest. In particular, the inhibitive role of polydispersity in the crystallization process of hard spheres is a contentious issue that deals with the value^{3,7–9} and even the very existence^{10,11} of a so-called terminal polydispersity σ_t , above which no crystallization can occur.

Moreover, one may wonder how polydispersity affects positional ordering if such ordering exists in just one or two dimensions, that is, in the case of smectic or columnar liquid crystals (sketched in Fig. 1). Driven by a gain in excluded volume entropy, these liquid crystal phases may be formed in concentrated suspensions of hard-body rod or plate-like particles^{12,13}. For dense systems of rods, computer simulation predicts a stable smectic phase up to a polydispersity in rod length of 18%, while for higher polydispersities the smectic phase is pre-empted by a columnar phase⁴. Accordingly, in experiments, almost monodisperse rod-like virus particles show a smectic phase⁶ while polydisperse solutions of DNA-rods show a columnar phase instead¹⁴.

The phase behaviour of hard plate-like particles, on the other hand, has received considerably less attention, largely because suitable experimental model systems have been developed only recently^{15,16}. One of these model systems, comprising fairly monodisperse platelets of low aspect ratio, exhibits an isotropic and columnar phase¹⁷; but, computer simulation for monodisperse platelets of sufficiently large aspect ratio predicts an isotropic (I),

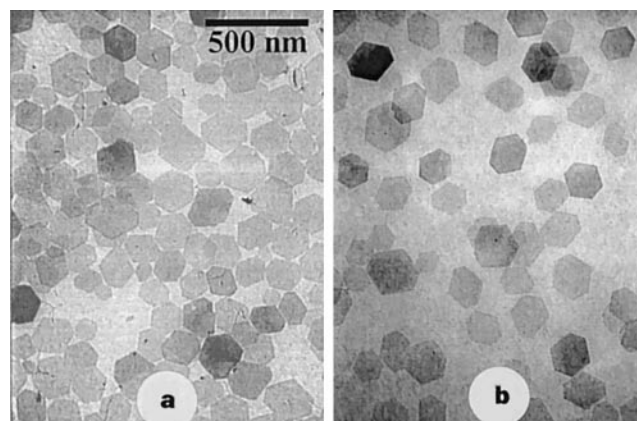


Figure 2 Transmission electron microscopy images. **a**, The parent suspension ($\sigma_D = 25\%$); and **b**, the fractionated suspension ($\sigma_D = 17\%$). The platelets are gibbsite (that is, $\text{Al}(\text{OH})_3$) colloids whose surface is grafted with a modified polyisobutylene ($M_w \approx 1,000$ g per mole). Owing to this steric stabilization layer, the particles interact through an approximately hard-core (that is, short-range repulsive) interaction potential when dispersed in apolar solvents (in this case, toluene)^{15,23}. The reduction of the particles' polydispersity in diameter to 17% is achieved by fractionation of part of the parent system, using a scheme which resembles the method of depletion-enhanced crystallization fractionation of emulsion droplets as described²⁴. The suspension is submitted to I–N phase separation, using ± 2 g l^{-1} added non-adsorbing polymer to enhance fractionation, yielding roughly 80% nematic phase in coexistence with 20% isotropic phase. The isotropic upper phase is subsequently removed. By dilution of the remaining nematic phase (using polymer solution) back to the I–N region and repeating this scheme twice, the suspension thus obtained has a lower polydispersity than the original system. The non-adsorbing polymer is subsequently removed by redispersing in polymer-free solvent after sedimentation. The number-average diameter $\langle D \rangle$ of the grafted platelets (Table 2) is based on the diameter of the core, determined from transmission electron microscopy (TEM) micrographs, plus twice the estimated thickness (4 nm) of the grafted polymer layer²⁵. We define the diameter of the hexagons by the diameter of a circle of equal area, with a relative standard deviation $\sigma_D = (\langle D^2 \rangle - \langle D \rangle^2)^{1/2} / \langle D \rangle$. The number-average thickness $\langle L \rangle$ of the plates (including the grafted polymer layer) is roughly 14 nm, with a standard deviation which is experimentally not readily accessible but probably lower than that in diameter²⁶.

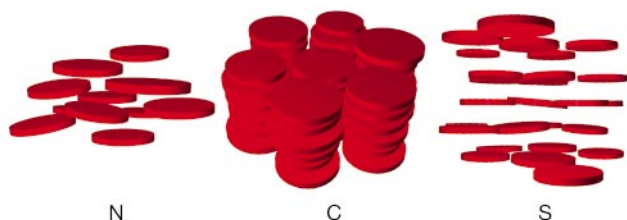


Figure 1 Structure of the three main classes of liquid crystals. The nematic phase (N), the columnar phase (C), and the smectic phase (S) are schematically depicted here for the case of plate-like particles. While each of these phases exhibits long-range orientational order, they differ by the positional correlations between the particles. In the nematic phase long-range positional order is absent. The columnar phase has a two-dimensional lattice of columns, which are constituted of liquid-like stacks of particles. The smectic phase is characterized by a one-dimensional periodic array of layers of particles.

a nematic (N), and a columnar (C) phase upon increasing the concentration of platelets^{13,18–21}. The isotropic to nematic (I–N) phase transition, which does not involve positional order (see Fig. 1), has recently been observed in experiments on suspensions of gibbsite platelets¹⁵. In the present study we explore the liquid crystal phase behaviour of these fairly polydisperse suspensions at higher densities, where simulations predict a N–C transition in the monodisperse case. We consider two systems which differ by the degree of polydispersity in diameter of the constituent platelets (17 and 25%, respectively, see Fig. 2). A dimension-based speculation on the terminal polydispersity for columnar ordering of plates, which can be considered as two-dimensional crystallization, would yield a value in the range 5–20%, falling between the values for three-dimensional crystallization of spheres and for one-dimensional smectic ordering of rods.

This raises the following questions. Do suspensions of plates, particularly in the studied range of polydispersity, show a nematic–columnar phase transition as predicted for monodisperse plates? Or do they, in order to circumvent the effect of polydispersity in diameter, form a smectic phase instead? What is the role and extent of particle size partitioning between coexisting phases?

The phase behaviour of the platelet suspensions as a function of the plate volume fraction is presented in Fig. 3. Isotropic–nematic phase coexistence is observed just below $\phi = 0.2$, yielding an isotropic upper phase and a birefringent nematic bottom phase. Macroscopic phase separation is complete within 12 hours. The width of the biphasic region $\Delta\phi_{IN}$ depends strongly on the polydispersity in diameter σ_D , $\Delta\phi_{IN}$ being twice as broad for the suspension with $\sigma_D = 25\%$ compared to that with $\sigma_D = 17\%$ (Fig. 4). This observation is consistent with the relation $\Delta\phi_{IN} \propto \sigma^2$ as found in computer simulations for polydisperse disks²⁰.

Upon increasing the plate volume fraction to roughly twice the I–N coexistence density, $\phi \approx 0.4$, both the suspension of 17 and 25% polydispersity enter a biphasic region where a nematic upper phase coexists with a more concentrated birefringent bottom phase. A columnar signature of the lower phase is suggested by unequivocal Bragg reflections when illuminated by white light (Fig. 3). The fact that these reflections appear for wavelengths of visible light demonstrates that the crystalline order pertains to a periodicity on a length scale of the plate diameter (characteristic of columnar ordering)

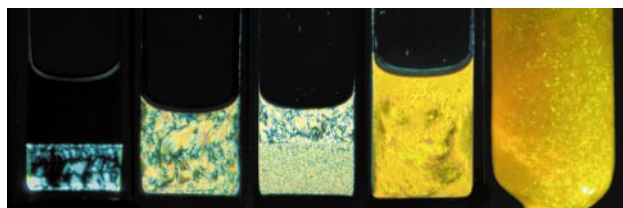


Figure 3 Tubes containing suspensions at varying concentrations. They are photographed between crossed polarizers to distinguish between isotropic (dark) and the birefringent nematic and columnar liquid crystalline phases. The suspensions depicted here comprise platelets with 17% polydispersity in diameter. From left to right, the concentration ranges from $\phi = 0.19$ (I + N), 0.28 (N), 0.41 (N + C), to 0.47 (C). The tube to the far right depicts the monophasic columnar sample at $\phi = 0.45$ as observed without polarizers but illuminated by white light. The colours of its Bragg reflections (visible as small bright spots) vary from yellow to green, as the angle between the incident light and viewing direction is in the range 50–70°. The particle volume fraction ϕ (which includes the solvent immobilized in the grafted polymer layer) is calculated as the mass concentration (determined by drying a known amount of dispersion to constant weight at 75 °C) divided by the effective mass density of the grafted particles. Following ref. 27, the latter is derived from the TEM particle dimensions, the estimated thickness of the stabilizing layer²⁵, the polymer mass fraction from elemental analysis and the mass density of gibbsite²⁸. This yields a mass density of about 1.3 g cm^{−3}, with an estimated error of 10% that is due to the uncertainty in polymer layer thickness.

rather than the much smaller plate thickness (as in smectic ordering). By applying Bragg's law to the angle of reflection measured for different wavelengths of light, we identify the characteristic spacing (of the (100) reflection, see small-angle X-ray scattering (SAXS) results) as 219 ± 5 nm and 214 ± 5 nm for the parent and fractionated suspensions, respectively. This corresponds to a typical distance between the centres of the columns of 253 ± 6 and 247 ± 6 nm. A comparison of the experimentally observed I–N and N–C transition densities to computer simulations for monodisperse hard disks¹³, as depicted in Fig. 4, shows that the transitions in the experiment are shifted to slightly lower densities. The difference in shape of the platelets studied (hexagonal in the experiment versus circular in the simulation) we expect to be a major contribution to this shift, as shown recently for the I–N transition²¹. Macroscopic phase separation in the N–C biphasic gap is complete within 2 weeks.

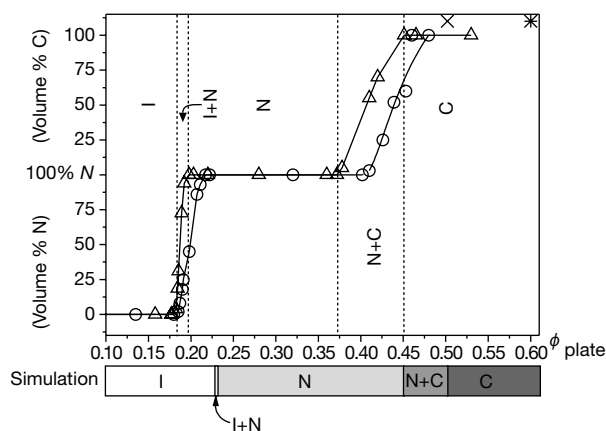


Figure 4 Phase diagram of the suspensions. The relative volume of the nematic and columnar phase are depicted after phase separation as a function of the platelet volume fraction ϕ . Results apply to $\sigma_D = 17\%$ (triangles) and $\sigma_D = 25\%$ (circles), respectively. The points marked by a cross and a star belong to the latter system but at densities beyond columnar stability, corresponding to curves **c** and **d**, respectively, in Fig. 5. The dotted lines indicate the boundaries of the coexistence regions of the suspension with $\sigma_D = 17\%$. Results from computer simulation¹³ for monodisperse hard disks, extrapolated to the current aspect ratio $\langle D \rangle / \langle L \rangle$ of roughly 13, are included for comparison.

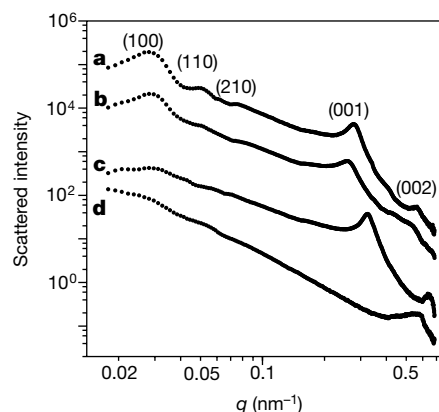


Figure 5 Small angle X-ray scattering patterns for samples varying in polydispersity and volume fraction. **a**, $\sigma_D = 17\%$ at $\phi = 0.45$, **b–d**, $\sigma_D = 25\%$ at $\phi = 0.45$, 0.5 and 0.6 respectively. Curves **a** and **b** correspond to a columnar phase, curve **c** to smectic-like ordering at densities above columnar stability, and curve **d** depicts the glassy state encountered upon increasing the density even further. Curves shown are radially averaged scattered intensities, in arbitrary units, as a function of the scattering vector $q = 4\pi/\lambda \sin(\theta/2)$, with θ the scattering angle. Experiments are performed at the DUBBLE beam line at the European Synchrotron Radiation Facility, France.

Table 1 q values of the scattering peaks in the SAXS pattern

Sample	σ_D (%)	ϕ	q (10^{-2} nm^{-1})				
			(100)	(110)	(210)	(001)	(002)
a	17	0.45	2.82	4.88	7.40	27.8	55.5
b	25	0.45	2.89	5.00	7.44	25.6	
c	25	0.5	(2.88)			32.3	64.4

SAXS experiments provide further proof for the columnar signature of the dense liquid crystal phase. Curves **a** and **b** in Fig. 5 correspond to the columnar phase in the case of 17 and 25% polydispersity, respectively. These curves show almost identical features. In the small q regime, where the spacing $d = 2\pi/q$ is of the order of the diameter of the plates, we can distinguish one major peak and two additional peaks. The q values of these three peaks (Table 1), whose q ratio is like $1 : \sqrt{3} : \sqrt{7}$, reveals that ordering in the plane of the plate diameters is hexagonal, with the peaks corresponding to the (100), (110) and (210) reflections. The (200) reflection seems to be relatively weak compared to its neighbouring (110) and (210) peaks, such that it is not (clearly) resolved in the scattering profile. From the q values of the (100), (110) and (210) peaks we obtain the typical distance between the centres of the columns to be 251 ± 4 and 258 ± 1 nm in the suspension with $\sigma_D = 25\%$ and 17% , respectively, in good agreement with the values found with light scattering. This distance is hence almost equal to $(1 + \sigma_D) \times D$ for both suspensions. The two peaks at much larger q correspond to 1 and 0.5 times a spacing of roughly the plate thickness, such that we identify them as (001) and (002) reflections. These peaks may therefore relate to (liquid-like) order between the plates along the z -axis of a column of plates. Although the observed scattering patterns are consistent with a columnar structure, they could also stem from a structure in which particles are hexagonally ordered in layers without lateral correlations between adjacent layers. These structures can be distinguished by orienting a single crystal in the X-ray beam with the beam parallel to the plate normals^{17,22}. If the structure is columnar, tilting the sample with respect to the beam will not result in a change in q values of the scattering peaks, similar to the behaviour of a three-dimensional crystal of spheres. Tilting a sample in a 0.2-mm flat capillary (approximating a single crystal) indeed does not lead to a significant ($>1\%$) shift in the q positions of the scattering peaks. Apart from the unlikely (though not excluded) possibility of a three-dimensional crystal structure, the tilting experiment thus confirms that the dense liquid crystal phase is a columnar phase.

A remaining question concerns the origin of the observed columnar stability in these systems where polydispersity must be an important factor. In analogy with computer simulations' predictions for crystallization of polydisperse hard spheres, stabilization of the ordered phase may emanate from fractionation, lowering the polydispersity in the ordered phase at the expense of the polydispersity in a coexisting disordered phase. In the present study we can determine the extent of fractionation experimentally, by examination (using transmission electron microscopy, TEM) of small samples of a coexisting nematic and columnar phase. Fractionation indeed gives rise to a reduction of the polydispersity in the columnar phase, as we find $\sigma_D = 18$ and 14% in the columnar phase

Table 2 Fractionation in the N-C coexistence region

System	Overall		N-phase		C-phase	
	$\langle D \rangle$ (nm)	σ_D (%)	$\langle D \rangle$ (nm)	σ_D (%)	$\langle D \rangle$ (nm)	σ_D (%)
I	198	25	196	26	200	18
II	212	17	209	18	220	14

The diameter distributions of the parent (I) and the pre-fractionated system (II), before and after N-C phase separation. Values are determined from TEM micrographs by measuring about 200 particles in each case.

of the two systems studied (Table 2). This indicates that at least the pre-fractionated system ($\sigma_D = 17\%$) should be able to enter a monophasic columnar state beyond the N-C coexistence region. In fact, the fully columnar state is observed for both suspensions. This demonstrates that even in the case of $\sigma_D = 25\%$ fractionation is not an absolute condition for columnar ordering, and that this polydispersity in diameter is therefore still below the terminal value. One may wonder, however, whether the columnar structure will persist upon increasing the volume fraction of a fully columnar sample, that is, if the spacing between the columns decreases such that the disruptive effect of polydispersity in diameter becomes more pronounced. For $\phi = 0.50$ (curve **c** in Fig. 5), the columnar (100), (110) and (210) peaks in a sample of $\sigma_D = 25\%$ become markedly suppressed, whereas, at the same time, the (001) and (002) peaks become more distinct. Bragg reflections for visible light (which pertain to the (100) reflection) almost completely disappear. We speculate that the suppression of the columnar peaks and the simultaneous structuring with a periodicity of the order of the plate thickness is indicative of a cross-over to smectic-like ordering. Unlike the columnar phase, a smectic phase is not sensitive to polydispersity in diameter, as ordering within the smectic layers is liquid-like. Instead, the smectic requires a low polydispersity in thickness. The relative stability of the nematic, columnar and smectic phases is therefore determined not only by the volume fraction of the suspension and the polydispersity in diameter, but also by polydispersity in thickness. Our observation that a system with σ_D as high as 25% forms a columnar phase, and its cross-over to a smectic-like structure at higher ϕ , are therefore probably connected with the fact that the platelets have a polydispersity in thickness too. Indeed, in the case of $\sigma_D = 17\%$, the columnar phase is still stable at $\phi = 0.53$ as demonstrated by strong Bragg reflections for visible light.

The stability of a columnar phase in these polydisperse suspensions of plates seems remarkable in the light of predictions for the terminal polydispersity for the crystal phase of hard spheres and smectic phase for hard rods, although in the latter only polydispersity in length is taken into account. Further insight into the stability of the nematic, columnar and smectic phase in systems of polydisperse plates therefore requires study by computer simulation, addressing the role of polydispersity in both diameter and thickness. □

Received 26 April; accepted 4 July 2000.

- Pusey, P. N. & van Megen, W. Phase behaviour of concentrated suspensions of nearly hard spheres. *Nature* **320**, 340–342 (1986).
- Blanco, A. *et al.* Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1.5 micrometers. *Nature* **405**, 437–440 (2000).
- Pusey, P. N. in *Les Houches, Session LI, Liquids, Freezing & Glass Transitions* Ch. 10 (North-Holland, Amsterdam, 1991).
- Bates, M. & Frenkel, D. Influence of polydispersity on the phase behavior of colloidal liquid crystals: A Monte Carlo study. *J. Chem. Phys.* **109**, 6193–6199 (1998).
- Bernal, J. D. & Fankuchen, I. X-ray and crystallographic studies of plant virus preparations. *J. Gen. Physiol.* **25**, 111–146 (1941).
- Dogic, Z. & Fraden, S. Smectic phase in a colloidal suspension of semiflexible virus particles. *Phys. Rev. Lett.* **78**, 2417–2420 (1997).
- Pusey, P. N. The effect of polydispersity on the crystallization of hard spherical colloids. *J. Phys. (Paris)* **48**, 709–712 (1987).
- Barrat, J. L. & Hansen, J. P. On the stability of polydisperse colloidal crystals. *J. Phys. (Paris)* **46**, 1547–1553 (1986).
- McRae, R. & Haymet, A. D. J. Freezing of polydisperse hard spheres. *J. Chem. Phys.* **88**, 1114–1125 (1988).
- Bolhuis, P. G. & Kofke, D. A. Monte Carlo study of freezing of polydisperse hard spheres. *Phys. Rev. E* **54**, 634–643 (1996).
- Kofke, D. A. & Bolhuis, P. G. Freezing of polydisperse hard spheres. *Phys. Rev. E* **59**, 618–622 (1999).
- Frenkel, D., Lekkerkerker, H. N. W. & Stroobants, A. Thermodynamic stability of a smectic phase in a system of hard rods. *Nature* **332**, 822–823 (1988).
- Veerman, J. A. C. & Frenkel, D. Phase behavior of disklike hard-core mesogens. *Phys. Rev. A* **45**, 5632–5648 (1992).
- Livolant, F., Levelut, A. M., Doucet, J. & Benoit, J. P. The concentrated DNA liquid crystalline phase is columnar hexatic. *Nature* **339**, 724–728 (1989).
- van der Kooij, F. M. & Lekkerkerker, H. N. W. Formation of nematic liquid crystals in suspensions of hard colloidal platelets. *J. Phys. Chem. B* **102**, 7829–7832 (1998).
- Brown, A. B. D., Clarke, S. M. & Rennie, A. R. Ordered phase of platelike particles in concentrated dispersions. *Langmuir* **14**, 3129–3132 (1998).

17. Brown, A. B. D., Ferrero, C., Narayanan, T. & Rennie, A. R. Phase separation and structure in a concentrated colloidal dispersion of uniform plates. *Eur. Phys. J. B* **11**, 481–489 (1999).
18. Onsager, L. The effect of shape on the interaction of colloidal particles. *Ann. NY Acad. Sci.* **51**, 627–659 (1949).
19. Frenkel, D. & Eppenga, R. Monte Carlo study of the isotropic-nematic transition in a fluid of thin hard disks. *Phys. Rev. Lett.* **52**, 1089–1092 (1982).
20. Bates, M. & Frenkel, D. Nematic-isotropic transition in polydisperse systems of infinitely thin hard platelets. *J. Chem. Phys.* **110**, 6553–6559 (1999).
21. Bates, M. Influence of particle shape on the nematic-isotropic transition of colloidal platelet systems. *J. Chem. Phys.* **111**, 1732–1736 (1999).
22. Guinier, A. *X-ray Diffraction in Crystals, Imperfect Crystals, & Amorphous Bodies* 169–172 (Dover, New York, 1994).
23. van der Kooij, F. M., Philipse, A. P. & Dhont, J. K. G. Sedimentation and diffusion in suspensions of sterically stabilized colloidal platelets. *Langmuir* **16**, 5317–5323 (2000).
24. Bibette, J. U. Depletion interactions and fractionated crystallisation for polydisperse emulsion purification. *J. Colloid Interface Sci.* **147**, 474–478 (1991).
25. Smits, C., Briels, W. J., Dhont, J. K. G. & Lekkerkerker, H. N. W. Influence of the stabilizing coating on the rate of crystallization of colloidal suspensions. *Prog. Colloid Polym. Sci.* **79**, 287–292 (1989).
26. van der Kooij, F. M. & Lekkerkerker, H. N. W. The liquid crystalline phase behaviour of a colloidal rod-plate mixture. *Phys. Rev. Lett.* **84**, 781–784 (2000).
27. van Bruggen, M. P. B., Dhont, J. K. G. & Lekkerkerker, H. N. W. Morphology and kinetics of the isotropic-nematic phase transition in dispersions of hard rods. *Macromolecules* **32**, 2256–2264 (1999).
28. Gitzen, W. H. *Alumina as a Ceramic Material* 31 (The American Ceramic Society, Columbus, 1970).

Acknowledgements

We thank A. R. Rennie for discussions concerning the columnar phase, I. Dolbnya and W. Bras for technical support at DUBBLE, and D. Frenkel for a critical reading of the manuscript. This work was supported by the Foundation for Fundamental Research on Matter (FOM) and the Netherlands Organization for the Advancement of Research (NWO).

Correspondence should be addressed to H.L. (e-mail: H.N.W.Lekkerkerker@chem.uu.nl).

Multiscale modelling of plastic flow localization in irradiated materials

Tomas Diaz de la Rubia*, Hussein M. Zbib†, Tariq A. Khraishi‡, Brian D. Wirth*, Max Victoria‡ & Maria Jose Caturla*

* Lawrence Livermore National Laboratory, 7000 East Avenue, L-353, Livermore, California 94550, USA

† Washington State University, School of Mechanical & Materials Engineering, Pullman, Washington 99164, USA

‡ EPFL-CRPP-Fusion Technology Materials, 5232 Villigen PSI, Switzerland

The irradiation of metals by energetic particles causes significant degradation of the mechanical properties^{1,2}, most notably an increased yield stress and decreased ductility, often accompanied by plastic flow localization. Such effects limit the lifetime of pressure vessels in nuclear power plants³, and constrain the choice of materials for fusion-based alternative energy sources⁴. Although these phenomena have been known for many years¹, the underlying fundamental mechanisms and their relation to the irradiation field have not been clearly demonstrated. Here we use three-dimensional multiscale simulations of irradiated metals to reveal the mechanisms underlying plastic flow localization in defect-free channels. We observe dislocation pinning by irradiation-induced clusters of defects, subsequent unpinning as defects are absorbed by the dislocations, and cross-slip of the latter as the stress is increased. The width of the plastic flow channels is limited by the interaction among opposing dislocation dipole segments and the remaining defect clusters.

The microstructure of irradiated materials evolves over a wide range of length and time scales, making radiation damage an inherently multiscale phenomenon. At the shortest scales

(nanometres and picoseconds), recoil-induced cascades of energetic atomic displacements give rise to a highly non-equilibrium concentration of point defects and point defect clusters⁵. Over macroscopic length and time scales these defects can migrate and alter the chemistry and microstructure, often inducing significant degradation of mechanical and other properties^{1,2}. In metals, the main features of neutron or ion beam irradiation-induced mechanical behaviour can be summarized as²: (1) a sharp increase in yield stress with irradiation dose; (2) the appearance of a yield point followed by a yield drop in f.c.c. metals; and (3) an instability that results in plastic flow localization within ‘dislocation channels’ and leads to loss of ductility and premature failure. An example of flow localization is shown in Fig. 1 (ref. 6). The transmission electron microscopy image shows a channel where all visible irradiation-induced point defect clusters are absent and where uniform large shear (about 100%) has taken place. The channels are 100 to 200 nm wide. Plastic flow localization is responsible for the observed loss of ductility.

Early theories of irradiation hardening focused on various source and dispersed barrier mechanisms^{7,8}. Recently^{9,10}, the analytical cascade-induced source hardening (CISH) model was proposed^{9,10}. This model uses insights from molecular dynamics simulations^{11–14} and accounts for some of the recent experimental observations. In the model, it is postulated that interstitial clusters produced in displacement cascades form glissile dislocation loops that migrate in one dimension by thermally activated glide, and decorate dislocations, thereby pinning them. However, while this model provides a rational explanation for the observed increase in yield stress, it does not account for plastic flow localization and the development of plastic instabilities.

Here we couple these experimental and atomistic simulation results^{11–18} to a three-dimensional dislocation dynamics (DD) simulation to investigate the relation between the irradiation field and mechanical behaviour. We consider two cases, Cu and Pd, which exhibit different damage morphologies under irradiation. Experiments have shown that vacancy stacking-fault tetrahedra (SFT) are the predominant defect type in low stacking fault energy Cu^{6,19}, whereas in high stacking fault energy Pd, self-interstitial atom Frank sessile loops constitute the majority of observed defects^{20,21}. Our DD simulation box is a cube 5 µm in size that contains an initial density of Frank–Read dislocation sources distributed at random on {111} planes. In our DD simulation^{22–24}, the plastic deformation of a single crystal is obtained by explicit accounting of the dislocation evolution history, that is, their motion and structure.

The motion and interaction of an ensemble of dislocations in a three-dimensional crystal is marched in time. Dislocations are discretized into straight-line segments of mixed character. The Peach–Koehler force F acting on a dislocation segment inside the computational cell is calculated from the stress fields that are caused by immediate neighbouring segments, all other dislocations segments, all defect clusters and the applied stress. The result is used to advance the dislocation segment based on a linear mobility model, $v_{gi} = M_{gi}F_{gi}$, where v_{gi} is the glide velocity of the dislocation segment, M_{gi} is the dislocation mobility, and F_{gi} is the glide component of the Peach–Koehler force minus the Peierl’s friction. On the basis of the history of dislocation motion, we obtain a measure for the macroscopic plastic strain rate. To ensure continuity of dislocation lines across the boundaries, we apply reflection boundary conditions²². Due to the long-range character of the dislocation stress field, long-range interactions are computed explicitly²².

In the simulation, segments that are on the verge of experiencing short-range interactions are identified. Based on a set of physical rules, such reactions may result in the formation of junctions, jogs, dipoles, and so on. The dislocations multiply by a variety of

mechanisms that may involve superjog collisions²⁵, standard Frank–Read source multiplication, and double cross slip. Cross slip is described as an activated process with probability P given by

$$P = \alpha \Omega_1 \delta t \exp\left(-\frac{\Delta W^* - \tau A}{kT}\right) \quad \text{for} \quad \Omega_1 = \frac{C_t \pi}{L}$$

where Ω_1 is the fundamental frequency of a vibrating dislocation segment of length L , C_t is the transverse sound velocity, δt is the time increment, α is a numerical parameter controlling the frequency of cross slip, ΔW^* is the activation energy for cross-slip²³, τ is the resolved shear stress, A is the area swept by the dislocation

segment that undergoes cross-slip, k is Boltzmann's constant and T is temperature.

The irradiation field is modelled by mapping into the DD box the spatial distribution of defect clusters (SFT and/or Frank sessile interstitial loops) that results from a particular irradiation fluence, temperature and particle energy. At room temperature, typical defect cluster radii are 0.5 to 3 nm and their density in the bulk, ρ_B , ranges from 10^{21} – 10^{24} m^{-3} , depending on irradiation fluence^{17,19,20}. In the simulation, this translated into 100,000 to 2 million defect clusters. Following the CISH model^{9,10}, Frank–Read sources decorated along the dislocation line with defect clusters are

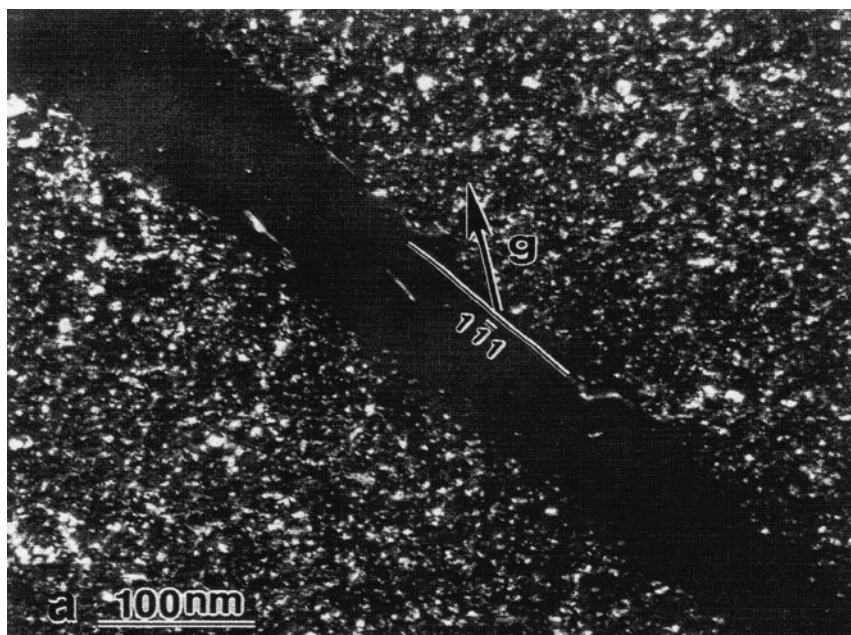


Figure 1 Defect-free channel in deformed irradiated Cu. Weak-beam dark-field transmission electron microscope image shows defect-free channels formed during

deformation of Cu irradiated with 600-MeV protons⁵.

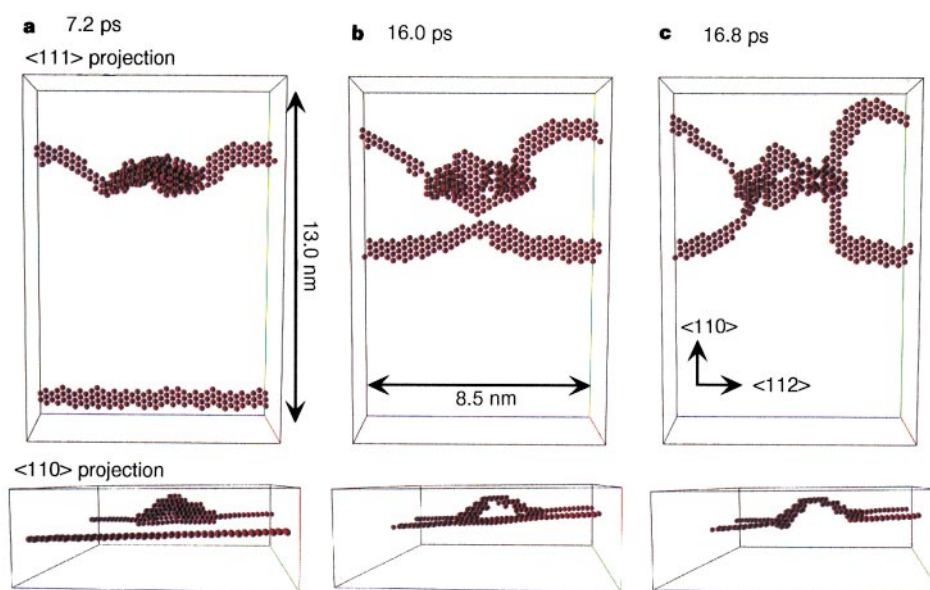


Figure 2 Molecular dynamics simulation of the interaction between an edge dislocation and an overlapping, truncated vacancy stacking-fault tetrahedron (SFT) in Cu. **a–c**, $\langle 111 \rangle$ (top) and $\langle 110 \rangle$ (bottom) projections of high-energy atoms show the motion of two Shockley partial dislocations on their $\langle 111 \rangle$ glide plane and their interaction with an overlapping vacancy SFT at **a** 7.2, **b** 16.0 and **c** 16.8 ps after application of a 300-MPa

shear stress. A triangular vacancy platelet that is bounded by $\langle 110 \rangle$ directions forms a single SFT. A triangular vacancy platelet bounded by two $\langle 110 \rangle$ and one $\langle 112 \rangle$ directions forms two partial SFTs, one above and one below the initial plane; we define this as an overlapping truncated SFT.

also introduced into the DD simulation as described below.

Dislocation segments interact with the irradiation-induced defect clusters. The elastic interaction of a moving dislocation with a defect cluster (SFT or loop) can be obtained from elasticity theory. However, elasticity calculations cannot by themselves determine the fate of the defect cluster as it interacts with the moving dislocation. To address this point, we have performed molecular dynamics simulations of dislocation–SFT interactions. Figure 2 shows the results of a simulation involving 933,000 atoms at 100 K in Cu. In this simulation, overlapping truncated SFTs (see Fig. 2 legend), such as formed by ageing displacement cascades¹⁷, with a spacing of 8.5 nm (along the dislocation line direction $\langle 112 \rangle$) are placed on the glide plane of a dissociated edge dislocation. The two partial dislocations then move under an applied shear stress of 300 MPa. Figure 2a shows that upon contact, the leading partial dislocation is initially pinned by the overlapping SFT. At a later time, Fig. 2b, the leading partial dislocation absorbs part of the overlapping SFT and climbs (clearly seen in $\langle 110 \rangle$ projection) as the trailing partial dislocation approaches. Finally, in Fig. 2c, we show that the trailing partial dislocation catches the leading partial dislocation at the location of the overlapping SFT, constricts and climbs by absorbing the remaining defect. Following absorption of the SFT, the dislocation continues to move but with reduced mobility as a result of the jogs that form on the constricted line segment. Although this simulation clearly shows defect absorption, the case is not unique and we are at present investigating the matrix of possible interactions.

For Frank sessile dislocation loops (relevant to Pd), the elastic interaction is computed explicitly²⁶. Here the Frank sessile loop is absorbed by the dislocation line if the loop ‘unfaults’ and becomes glissile. We can calculate the energy difference²⁷, ΔE , between the faulted loop state and the unfaulted state (that is, a perfect loop), taking into account the interaction energy between the dislocation and the sweeping Shockley partials. When $\Delta E > 0$, the critical radius for loop unfaulting decreases and the loop may transform from the faulted sessile to the unfaulted glissile configuration. When the critical radius for unfaulting is reached, the loop gets absorbed at the core of the dislocation and the dislocation is able to move. Recently, atomistic simulations for interstitial loops consistent with this model have been carried out²⁸.

Figure 3 shows the stress–strain curves obtained with our DD

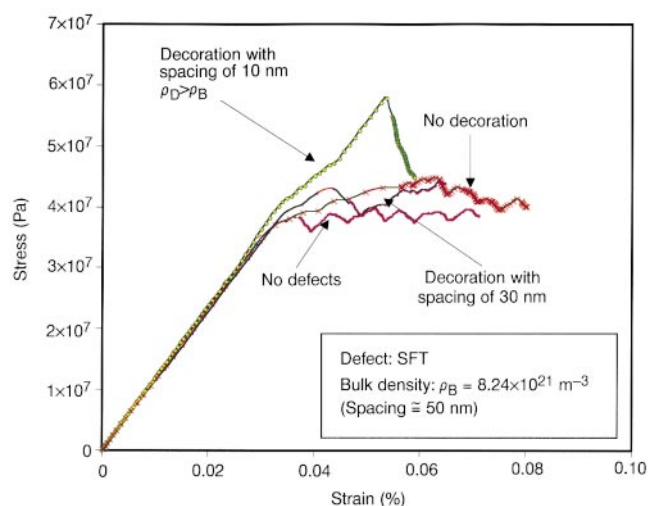


Figure 3 Dislocation dynamics (DD) results of stress–strain curves. Shown are the stress–strain curves obtained with our DD simulations for Cu irradiated to $\rho_B = 8.24 \times 10^{21} \text{ m}^{-3}$. Without irradiation the system yields at about 37 MPa. When the defect density along the line (ρ_D) is such that $\rho_D \gg \rho_B$, a clear yield point followed by a yield drop can be observed. However, when the average spacing between defects along the dislocation line is 30 nm, which is comparable to that in the bulk, the system yields at a slightly higher stress and only a very diffuse yield drop is apparent.

simulations for Cu irradiated to $\rho_B = 8.24 \times 10^{21} \text{ m}^{-3}$, which considering a uniform defect distribution, corresponds to a SFT spacing of 50 nm. The initial network dislocation density is $\rho_N = 10^{12} \text{ m}^{-2}$. Without irradiation the system yields at about 37 MPa, and work hardening starts at the early stage of deformation as a result of dislocation multiplication and forest interactions. When irradiation-induced defects are present, two behaviours are seen. When the defect density along the line (ρ_D) is such that the average spacing is only 10 nm, that is, $\rho_D \gg \rho_B$, a clear yield point followed by a yield drop can be observed. However, when the average spacing between defects along the dislocation line is 30 nm, which is comparable to that in the bulk, the system yields at a slightly higher stress and only a very diffuse yield drop is apparent. This is in good agreement with experiments⁵ which show that the observed yield drop disappears at elevated doses where $\rho_D \approx \rho_B$. The dose dependence of the yield stress obtained in our DD simulations of irradiated Pd (for more details see ref. 29), where all the irradiation-induced defect clusters are Frank sessile dislocation loops, has a slope of 0.35, which is in excellent agreement with experimental observations²¹.

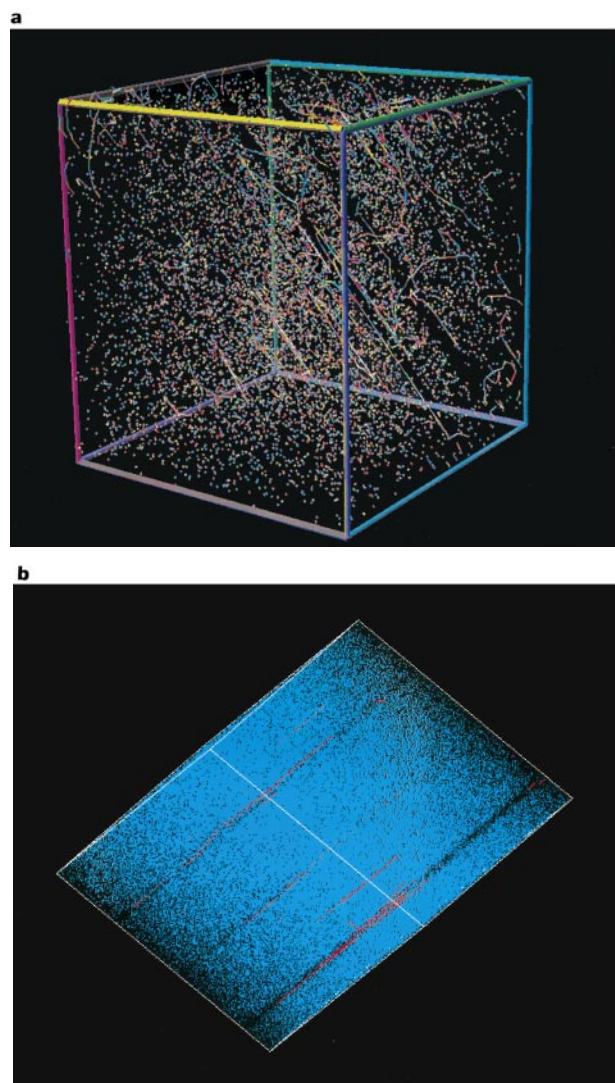


Figure 4 Dislocation dynamics results of channel formation and flow localization. **a**, Results of DD simulation in Cu irradiated to $\rho_B = 8.24 \times 10^{21} \text{ m}^{-3}$ with an initial network dislocation density of $\rho_N = 10^{12} \text{ m}^{-2}$. Shown is the 10- μm side computational cell containing the dislocations and point defects with defect-free channels. **b**, Two-dimensional projection of **a**.

Figure 4a and its projection into two dimensions (Fig. 4b) show the results of our simulations in Cu where cross-slip/double cross-slip takes place and irradiation-induced SFT defects are absorbed by sweeping dislocations based on the criterion discussed above. Figure 4 shows localization of plastic flow in defect-free channels. The channel width is approximately 200 nm with a channel spacing of 1,000 nm in agreement with experimental observations¹. Examination of the simulation results reveals that the mechanism of channel formation can be described as follows. As dislocations propagate from a Frank–Read source, a screw dislocation segment may be pinned by the defects and eventually cross slip from its primary plane as the stress is increased. This is followed by double cross-slip of the segment, creating a new dislocation source on a parallel plane. This process continues with dislocation segments cross-slipping progressively from one glide plane to another, resulting in a set of parallel planes with active dislocations (the spacing between each pair of planes varies between 25 and 50 nm). Eventually, spreading of the band is limited by the interaction between segments of opposite sign that, in concert with the defects, exerts a backstress on those segments attempting to cross-slip. However, the process is continuous with dislocations being emitted from the same original single source, cross-slipping to adjacent planes and so on, resulting in a large local shear strain (about 100%) in the band. The spacing between the channels is a stochastic variable determined by the spatial distribution of dislocation sources. □

Received 23 February; accepted 30 June 2000.

1. Eyre, B. L. & Barlett, A. F. An electron microscope study of neutron irradiation damage on α -iron. *Phil. Mag.* **12**, 261–271 (1965).
2. Victoria, M. *et al.* The microstructure and associated tensile properties of irradiated fcc and bcc metals. *J. Nucl. Mater.* **276**, 114–122 (2000).
3. Odette, G. R. & Lucas, G. E. Recent progress in understanding reactor pressure vessel steel embrittlement. *Radiat. Effects Defects Solids* **144**, 189–231 (1998).
4. Bloom, E. E. The challenge of developing structural materials for fusion power systems. *J. Nucl. Mater.* **258–263**, 7–17 (1998).
5. Averbach, R. S. & Diaz de la Rubia, T. in *Solid State Physics* Vol. 51 (eds Spaepen, F. *et al.* & Ehrenreich, H.) 281–402 (Academic, New York, 1998).
6. Dai, Y. & Victoria, M. Defect cluster structure and tensile properties of Cu single crystals irradiated with 600 MeV protons. *MRS Symp. Proc.* **439**, 319–324 (1997).
7. Friedel, J. *Dislocations* (Pergamon, Oxford, 1964).
8. Bement, A. L. Fundamental materials problems in nuclear reactors. In *Proc. 2nd Int. Conf. Strength of Metals Alloys* Vol. II, 693–728 (The American Society for Metals, 1970).
9. Trinkaus, H., Singh, B. N. & Foreman, A. J. E. Segregation of cascade induced interstitial loops at dislocations: possible effect on initiation of plastic deformation. *J. Nucl. Mater.* **251**, 172–187 (1997).
10. Singh, B. N., Foreman, A. J. E. & Trinkaus, H. Radiation hardening revisited: role of intracascade clustering. *J. Nucl. Mater.* **249**, 103–115 (1997).
11. Diaz de la Rubia, T. & Guinan, M. W. New mechanism of defect production in metals: A molecular-dynamics study of interstitial-dislocation-loop formation in high-energy displacement cascades. *Phys. Rev. Lett.* **66**, 2766–2769 (1991).
12. Foreman, A. J. E., Phythian, W. J. & English, C. A. The molecular dynamics simulation of irradiation damage cascades in copper using a many-body potential. *Phil. Mag.* **66**, 671–695 (1992).
13. Wirth, B. D., Odette, G. R., Maroudas, D. & Lucas, G. E. *J. Nucl. Mater.* **244**, 185–194 (1997).
14. Osotsky, Y. N., Bacon, D. J. & Serra, A. Thermally activated glide of small dislocation loops in metals. *Phil. Mag. Lett.* **79**, 273–282 (1999).
15. Wirth, B. D., Bulatov, V. & Diaz de la Rubia, T. Atomistic simulation of stacking fault tetrahedra formation in Cu. *Proc. Int. Conf. Fusion Reactor Materials ICFRM-9*, *J. Nucl. Mater.* (in the press).
16. Heinisch, H. L. Computer simulation of high energy displacement cascades. *Radiat. Effects Defects Solids* **113**, 53–73 (1990).
17. Caturla, M. J. *et al.* Comparative study of radiation damage accumulation in Cu and Fe. *J. Nucl. Mater.* **276**, 13–21 (2000).
18. Soneda, N. & Diaz de la Rubia, T. Defect production, annealing kinetics and damage evolution in alpha-Fe: an atomic-scale computer simulation. *Phil. Mag.* **78**, 995–1019 (1998).
19. Singh, B. N. & Zinkle, S. J. Defect accumulation in pure fcc metals in the transient regime: a review. *J. Nucl. Mater.* **206**, 212–229 (1993).
20. Baluc, N., Dai, Y. & Victoria, M. Plasticity and microstructure of irradiated Pd. *MRS Symp. Proc.* **540**, 555–560 (1999).
21. Baluc, N., Bailat, C., Luppo, M.-I., Schaublin, R. & Victoria, M. Comparison of the microstructure and tensile behavior irradiated fcc and bcc metals. *MRS Symp. Proc.* **540**, 539–548 (1999).
22. Zbib, H. M., Rhee, M. & Hirth, J. P. On plastic deformation and the dynamics of 3D dislocations. *Int. J. Mech. Sci.* **40**, 113–127 (1998).
23. Rhee, M., Zbib, H. M., Hirth, J. P., Huang, H. & Diaz de la Rubia, T. Models for long/short range interactions in 3D dislocation simulation. *Model. Simul. Mater. Sci. Eng.* **6**, 467–492 (1998).
24. Zbib, H. M., Diaz de la Rubia, T., Rhee, M. & Hirth, J. P. 3D dislocation dynamics: stress-strain behavior and hardening mechanisms in fcc and bcc metals. *J. Nucl. Mater.* **276**, 154–165 (2000).
25. Rhee, M., Lassila, D. H., Bulatov, V. V., Hshuing, L. & Diaz de la Rubia, T. Dislocation multiplication and the early stages of plastic deformation in bcc molybdenum: a dislocation dynamics numerical simulation. *Phys. Rev. Lett.* (submitted).

26. Khraishi, T. A., Zbib, H. M., Hirth, J. P. & Diaz de la Rubia, T. The stress field of a general circular Volterra dislocation loop: analytical and numerical approaches. *Phil. Mag. Lett.* **80**, 95–105 (2000).
27. Ghoniem, N. M., Singh, B. N., Sun, L. Z. & Diaz de la Rubia, T. Interaction and accumulation of glissile defect clusters near dislocations. *J. Nucl. Mater.* **276**, 166–177 (2000).
28. Rodney, D. & Martin, G. Dislocation pinning by small interstitial loops: a molecular dynamics study. *Phys. Rev. Lett.* **82**, 3272–3275 (1999).
29. Khraishi, T. A., Zbib, H. M., Diaz de la Rubia, T. & Victoria, M. Localized deformation and hardening in irradiated metals: 3D discrete dislocation dynamics simulations. *Acta Metall. Mater.* (submitted).

Acknowledgements

This work was performed in part under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory. Work by M.V. was performed under a contract of the Swiss National Research Fund.

Correspondence and requests for materials should be addressed to T.D.R. (e-mail: delarubia@llnl.gov).

A stable argon compound

Leonid Khriachtchev, Mika Pettersson, Nino Runeberg, Jan Lundell & Markku Räsänen

Department of Chemistry, PO Box 55 (A.I. Virtasen aukio 1), FIN-00014 University of Helsinki, Finland

The noble gases have a particularly stable electronic configuration, comprising fully filled *s* and *p* valence orbitals. This makes these elements relatively non-reactive, and they exist at room temperature as monatomic gases. Pauling predicted¹ in 1933 that the heavier noble gases, whose valence electrons are screened by core electrons and thus less strongly bound, could form stable molecules. This prediction was verified in 1962 by the preparation of xenon hexafluoroplatinate, XePtF₆, the first compound to contain a noble-gas atom^{2,3}. Since then, a range of different compounds containing radon, xenon and krypton have been theoretically anticipated and prepared^{4–8}. Although the lighter noble gases neon, helium and argon are also expected to be reactive under suitable conditions^{9,10}, they remain the last three long-lived elements of the periodic table for which no stable compound is known. Here we report that the photolysis of hydrogen fluoride in a solid argon matrix leads to the formation of argon fluorohydride (HArF), which we have identified by probing the shift in the position of vibrational bands on isotopic substitution using infrared spectroscopy. Extensive *ab initio* calculations indicate that HArF is intrinsically stable, owing to significant ionic and covalent contributions to its bonding, thus confirming computational predictions^{11–13} that argon should form a stable hydride species with properties similar to those of the analogous xenon and krypton compounds reported before^{14–18}.

We prepared hydrogen fluoride in an argon matrix by passing argon over an HF–pyridine polymer (Fluka) at room temperature, and condensing the mixture onto a CsI substrate kept at 7.5 K. The argon used was either ⁴⁰Ar (100%; AGA) or ³⁶Ar (99.5% isotopically enriched; ICON Services). By optimizing the experimental conditions, fairly monomeric samples of HF were obtained in the matrix (infrared bands corresponding to H–F stretching vibrations were found at $\nu_{\text{HF}} = 3,962$ and $3,954 \text{ cm}^{-1}$; $\nu_{\text{DF}} = 2,895 \text{ cm}^{-1}$), as evidenced by comparison with previous HF/Ar matrix-isolation studies¹⁹. A very high degree of deuteration (> 90%) was achieved by passing the gaseous HF/Ar mixture through a volume containing small amounts of liquid deuterated sulphuric acid.

Photolysis of the HF was performed by illuminating the HF/Ar matrix (through a MgF₂ window) with a Kr vacuum-ultraviolet continuum discharge lamp (Ophos), which emitted in the spectral region of 127–160 nm wavelength. Typically, only a fraction (< 30%) of HF dissociated when this light source was used, even

after extended periods of photolysis; this most probably indicates self-limitation of the photolysis due to formation of absorbing species²⁰, possibly involving charge-transfer excitation of impurity oxygen atoms.

After photolysis of the HF, three previously unknown bands appeared in the infrared spectra at about 1,969.5, 687.0 and 435.7 cm⁻¹ in an ⁴⁰Ar environment; the highest-frequency band had three components, assigned to different matrix sites. The intensities of these new absorptions grew proportionally on annealing the sample at about 18 K, and the absorber was found to be stable in our experiments at temperatures below about 27 K. A similar set of new absorptions was obtained in photolysed and annealed DF/⁴⁰Ar matrixes, the positions now being 1,466.3, 513.0 and 435.3 cm⁻¹. In HF/³⁶Ar samples, the three absorptions shifted to higher frequencies, these shifts being approximately +2, +2 and +7 cm⁻¹ from the values given above for HF in ⁴⁰Ar matrix.

The infrared absorption bands obtained by photolysis and annealing of HF/⁴⁰Ar, HF/³⁶Ar and DF/⁴⁰Ar matrixes are shown in Fig. 1. The proportional growth of the three new bands on annealing the sample at different temperatures, as well as their synchronous decrease on ultraviolet irradiation (see below) prove that they belong to the same species. We emphasize that the three infrared absorptions under discussion are completely absent in photolysed and annealed matrices of 'pure' Ar—containing only minor amounts of normal impurities like O₂, N₂, CO₂ and H₂O—showing the participation of F atoms in the absorbing species. Based on deuteration shifts, the two highest-frequency absorptions are predominantly due to hydrogen atom motion. The positions of these absorptions are obviously too low to be due to H–F, H–O, H–N or H–C bonds. The van der Waals complexation of Ar atoms to molecules typically results in vibrational shifts of a few wavenumbers; the greatest shift observed to date (62 cm⁻¹, ~ 4%) being due to an exceptionally strong interaction between Ar and BeO (refs 21, 22). Strong van der Waals interactions between Ar and CuX or AgX

(X = F, Cl, Br) have recently been observed^{23,24}. Furthermore, the 7.2-cm⁻¹ shift of the lowest-wavenumber band on ⁴⁰Ar → ³⁶Ar isotopic substitution shows the presence of Ar chemically bound to a heavier element, which, based on elementary vibrational treatment, is most probably fluorine. On this analysis we assign the three absorptions that we observe to $\nu(\text{H}-\text{Ar})$, $\delta(\text{H}-\text{Ar}-\text{F})$ and $\nu(\text{Ar}-\text{F})$ of HArF, isolated in solid Ar. (Here δ indicates a bending vibration.)

HArF photobleaches under ultraviolet irradiation, in a similar fashion to the previously characterized Kr- or Xe-containing molecules^{14–17}. The threshold for photodecomposition of HArF is in the region of 350–400 nm. The infrared absorptions of HArF decrease at temperatures above 27 K, when the different photoproducts start to move throughout the matrix; this motion is particularly indicated by the formation of O₃, HOO and FOO. We believe that the mobilization of individual atoms originating from photodissociated HF and matrix impurities is responsible for the decrease of HArF at temperatures above 27 K in solid Ar, due to secondary reactions. Owing to the presence of these thermally activated secondary reactions, the question of the intrinsic thermal stability of HArF in solid Ar remains unanswered for the time being.

The characteristic triplet structure of $\nu(\text{H}-\text{Ar})$ appears in DF samples, with a maximum at about 1,466.3 cm⁻¹, resulting in a relative deuteration shift (that is, $\nu(\text{H}-\text{Ar})/\nu(\text{D}-\text{Ar})$ of 1.343. It is worth comparing this value with the deuteration shifts obtained for similar Xe- and Kr-containing hydrides^{14–16}. For the strongly bound Xe compounds HXeCN and HXeCl, the shifts are 1.382 and 1.376, respectively. On the other hand, the corresponding numbers for the weakly bound compounds HXeI and HXeSH are 1.336 and 1.343, respectively. The changes in the deuteration shifts display the effect of increasing anharmonicity with the weakening bond between hydrogen and the noble gas. The shift of HArF in Ar is close to the value found for the more weakly bound hydrides containing heavier noble gases.

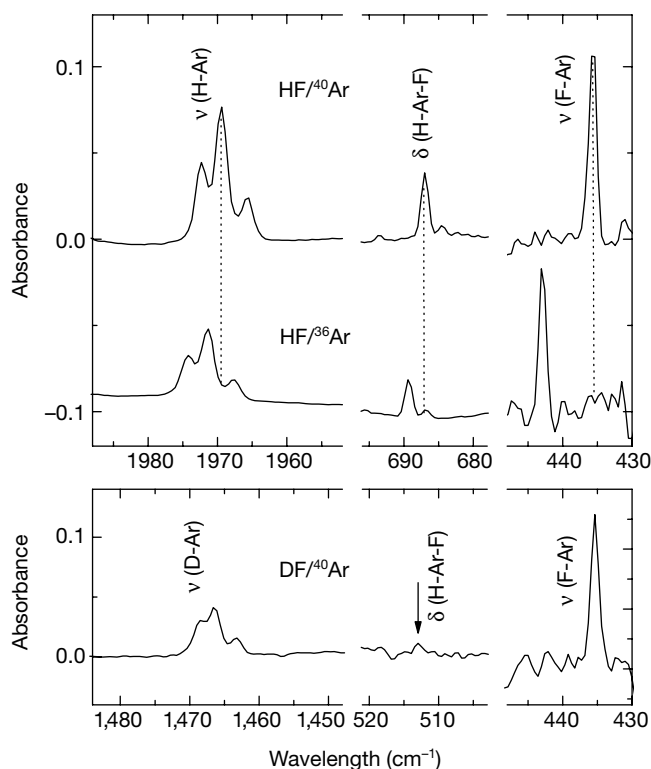


Figure 1 The infrared absorptions of HArF in solid Ar at 7.5 K. This figure shows the effect of ⁴⁰Ar/³⁶Ar and H/D isotopic substitutions. The increased noise in the low-frequency region is due to the low sensitivity of the detector in this region. Note the different scales in

the plot. The spectra were recorded at 1 cm⁻¹ resolution on a Nicolet 60 SX Fourier-transform infrared (FTIR) spectrometer.

We used molecular-orbital calculations to investigate the molecular properties of HARF: perturbation-theory (MP2) and coupled-cluster (CCSD(T)) calculations were performed using the 'aug-cc-pC5Z' and 'aug-cc-pVTZ' basis sets. All the approaches that we used indicate that the molecule is linear; the CCSD(T)/aug-cc-pV5Z computations yielded H–Ar and Ar–F bond distances of 133 and 197 pm, respectively. The NBO (natural bond orbital) charge distribution obtained at the MP2/aug-cc-pVTZ level of theory indicates a strong ionic character of the molecule, with a positive partial charge on argon (+0.54) and a negative partial charge on fluorine (–0.76). The calculated dipole moment at the CCSD(T)/aug-cc-pV5Z level is 6.51 D. In analogy to the other noble-gas hydrides^{14–16} this is an indication of the strong (HAr)⁺F[–] charge transfer in the HARF electronic structure. This shows that HARF differs completely from the predicted salts containing the strongly bound (ArF)⁺ unit^{9,10}. In order to get an idea of the relative proportions of ionic and covalent contributions in the bonding of HARF, we compare the computed bond lengths and experimental frequencies with published data. The gas-phase harmonic frequency of HAR⁺ is 2,710.9 cm^{–1} and the bond distance 128 pm (ref. 25); comparison with the corresponding values for HARF indicates that the complete ionic limit (HAr)⁺F[–] is not dominating in the molecule. On the other hand, for Ar⁺F[–] in the ²Σ⁺ state the calculated bond length is 240 pm and the vibrational frequency is 394 cm^{–1} (ref. 26), and the observed frequency in a neon matrix is 415.5 cm^{–1} (ref. 27). These values differ from the present values for HARF—197 pm and 435.7 cm^{–1}—and suggest the presence of covalent contribution in the bonding. Furthermore, the bond length and vibrational frequency of (ArF)⁺ (163.7 pm and 750 cm^{–1})¹⁰ differ greatly from the corresponding values for HARF.

At the CCSD(T)/aug-cc-pV5Z level, the HARF molecule is 5.87 eV higher in energy than the Ar + HF configuration. The calculated minimum dissociation barrier is along the H–Ar stretching coordinate, and its value is about 0.35 eV. The bending coordinate is connected to a much higher barrier than is the H–Ar stretching, and (according to our computations) it is above 1 eV for HARF. In addition, solvation of a strong dipole in a dielectric host probably provides further stabilization of HARF with respect to its dissociation limit, but this effect is not expected to be a determining factor in the existence of HARF. We note here that our computations refer to the free molecule.

The experimental and computational results for HARF (⁴⁰Ar and ³⁶Ar) and DARF are compared in Table 1. The CCSD(T)/aug-cc-pVQZ calculations predict the following three vibrations for HARF: Ar–H stretching (2,097 cm^{–1}), Ar–H bending (729 cm^{–1}), and Ar–F stretching (480 cm^{–1}). These vibrational frequencies differ only nominally from those obtained with the smaller VTZ basis. The lower-level MP2/aug-cc-pVTZ calculations give similar results to the CCSD(T) calculations except for the Ar–H stretch, which is now at 2,313 cm^{–1}. The experimental and calculated infrared intensities are very similar, and the calculations predict a considerable intensity decrease for the bending mode on deuteration. For the

Kr- and Xe-containing molecules, the hydrogen/noble-gas stretching mode has been noted to be highly anharmonic¹⁴. Therefore, anharmonic corrections were derived for HARF from the MP2-calculations²⁸. Indeed, at the MP2/aug-cc-pVTZ level the anharmonic effect is about 210 cm^{–1} for the Ar–H stretching vibration. For the bending, and the heavy-atom stretching, modes the corrections are smaller, 36 and 8 cm^{–1}, respectively. When the anharmonicity is accounted for, the agreement between the calculated and experimental wavenumbers improves considerably.

The ³⁶Ar–⁴⁰Ar shift is especially sensitive to the nature of the Ar–F bond; the calculated shift is practically the same as that experimentally observed. We note that our determination of the stretching frequency of the Ar–F bond is clear evidence for the existence of that bond; to our knowledge, this is the first infrared determination of the noble-gas/heavy-atom stretching frequency in the family of noble-gas-containing hydrides. □

Received 6 April; accepted 19 June 2000.

- Pauling, L. The formulas of antimonic acid and the antimonates. *J. Am. Chem. Soc.* **55**, 1895–1900 (1933).
- Bartlett, N. Xenon hexafluoroplatinate(V) Xe⁺[PtF₆[–]]. *Proc. Chem. Soc.* 218 (1962).
- Graham, L., Graedel, O., Jha, N. K. & Bartlett, N. Concerning the nature of XePtF₆. *Coord. Chem. Rev.* **197**, 321–334 (2000).
- Holloway, J. H. & Hope, E. G. Recent advances in noble-gas chemistry. *Adv. Inorg. Chem.* **46**, 51–100 (1999).
- Nelson, L. Y. & Pimentel, G. C. Infrared detection of xenon dichloride. *Inorg. Chem.* **6**, 1758–1759 (1967).
- Turner, J. J. & Pimentel, G. C. Krypton fluoride: preparation by the matrix isolation technique. *Science* **140**, 974–975 (1963).
- Bondybey, V. E. *Matrix Isolation Search for Transient Species*. Thesis, Univ. California, Berkeley (1971).
- Stein, L. Removal of xenon and radon from contaminated atmospheres with dioxygenyl hexafluoroantimonate, O₂SbF₆. *Nature* **243**, 30–32 (1973).
- Frenking, G. & Cremer, D. The chemistry of the noble gas elements helium, neon and argon. Facts and theoretical predictions. *Struct. Bonding* **73**, 17–95 (1990).
- Frenking, G., Koch, W., Cremer, D., Gauss, J. & Liebman, J. F. Neon and argon bonding in first-row cations NeX⁺ and ArX⁺ (X = Li–Ne). *J. Phys. Chem.* **93**, 3410–3418 (1989).
- Petterson, M., Lundell, J. & Räsänen, M. Neutral rare gas containing charge-transfer molecules in solid matrices: HXeCl, HXeBr, HXeI and HKrCl in Kr and Xe. *J. Chem. Phys.* **102**, 6423–6431 (1995).
- Runeberg, N., Petterson, M., Khriachtchev, L., Lundell, J. & Räsänen, M. A theoretical study of HARF: an observed neutral argon compound. *J. Chem. Phys.* (submitted).
- Wong, M. W. Prediction of a metastable helium compound: H He F. *J. Am. Chem. Soc.* **122**, 6289–6290 (2000).
- Petterson, M., Lundell, J. & Räsänen, M. New rare-gas-containing neutral molecules. *Eur. J. Inorg. Chem.* 729–737 (1999).
- Petterson, M., Khriachtchev, L., Lundell, J. & Räsänen, M. A chemical compound formed from water and xenon: HXeOH. *J. Am. Chem. Soc.* **121**, 11904–11905 (1999).
- Petterson, M., Khriachtchev, L., Lundell, J., Jolkonen, S. & Räsänen, M. Photochemistry of HNCO in solid xenon: Photoinduced and thermally activated formation of HXeNCO. *J. Phys. Chem. A* **104**, 3579–3583 (2000).
- Petterson, M., Nieminen, J., Khriachtchev, L. & Räsänen, M. The mechanism of formation and IR-induced decomposition of HXeI in solid Xe. *J. Chem. Phys.* **107**, 8423–8431 (1997).
- Lorenz, M., Räsänen, M. & Bondybey, V. E. Neutral xenon hydrides in solid neon and their intrinsic stability. *J. Phys. Chem. A* **104**, 3770–3774 (2000).
- Hunt, R. D. & Andrews, L. Photolysis of hydrogen fluoride in solid argon. Matrix infrared spectra of (HF)₂, (HF)(DF), and (DF)₂. *J. Chem. Phys.* **82**, 4442–4448 (1985).
- Khriachtchev, L., Petterson, M. & Räsänen, M. On self-limitation of UV photolysis in rare-gas solids and some of its consequences for matrix studies. *Chem. Phys. Lett.* **288**, 727–733 (1998).
- Thompson, C. A. & Andrews, L. Noble gas complexes with BeO: infrared spectra of Ng–BeO (Ng = Ar, Kr, Xe). *J. Am. Chem. Soc.* **116**, 423–424 (1994).
- Frenking, G., Koch, W., Gauss, J. & Cremer, D. Stabilities and nature of the attractive interactions in HeBeO, NeBeO, and ArBeO and a comparison with analogues NGLiF, NGBN, and NGLiH (NG = He, Ar). A theoretical investigation. *J. Am. Chem. Soc.* **110**, 8007–8016 (1988).
- Evans, C. J. & Gerry, M. C. L. Noble gas-metal chemical bonding? The microwave spectra, structures, and hyperfine constants of Ar–CuX (X = F, Cl, Br). *J. Chem. Phys.* **112**, 9363–9374 (2000).
- Evans, C. J. & Gerry, M. C. L. The microwave spectra and structures of Ar–AgX (X = F, Cl, Br). *J. Chem. Phys.* **112**, 1321–1329 (2000).
- Johns, J. W. C. Spectra of the protonated rare gases. *J. Mol. Spectrosc.* **106**, 124–133 (1984).
- Dunning, T. H. Jr & Hay, P. J. The covalent and ionic states of the rare gas monofluorides. *J. Chem. Phys.* **69**, 134–149 (1978).
- Bressler, C., Lawrence, W. G. & Schwentner, N. Spectroscopy of argon fluoride and krypton fluoride exciplexes in rare gas matrices. *J. Chem. Phys.* **105**, 10178–10188 (1996).
- Chaban, G. M., Jung, J. O. & Gerber, R. B. Ab initio calculations of anharmonic vibrational states of polyatomic systems: Electronic structure combined with vibrational self-consistent field. *J. Chem. Phys.* **111**, 1823–1829 (1999).

Acknowledgements

We thank P. Pykkö for discussions. This work was supported by The Academy of Finland.

Correspondence and requests for materials should be addressed to M.R. (e-mail: markku.rasanen@helsinki.fi).

Table 1 Experimental and calculated vibrations of HARF

	ν (Ar–H) (cm ^{–1})	δ (H–Ar–F) (cm ^{–1})	ν (Ar–F) (cm ^{–1})
H– ⁴⁰ Ar–F			
Experimental	1,969.5*	687.0	435.7
MP2/aug-cc-pVTZ (harm.)	2,313	749	482
MP2/aug-cc-pVTZ (anharm.)	2,103	713	474
CCSD(T)/aug-cc-pVTZ (harm.)	2,053	725	474
CCSD(T)/aug-cc-pVQZ (harm.)	2,097	729	480
D– ⁴⁰ Ar–F			
Experimental	1,466.3*	513.0	435.3
CCSD(T)/aug-cc-pVQZ (harm.)	1,500	538	479
H– ³⁶ Ar–F			
Experimental	1,971.3*	689.3	442.9
CCSD(T)/aug-cc-pVQZ (harm.)	2,100	731	488

* The strongest component of the site-split absorption.

Changes in Greenland ice sheet elevation attributed primarily to snow accumulation variability

J. R. McConnell*, R. J. Arthern†, E. Mosley-Thompson‡, C. H. Davis§, R. C. Bales||, R. Thomas¶, J. F. Burkhart|| & J. D. Kyne#

* Desert Research Institute, University & Community College System of Nevada, Reno, Nevada 89512, USA

† Applied Physics Laboratory, University of Washington, Seattle, Washington 98105, USA

‡ Byrd Polar Research Centre & Department of Geography, The Ohio State University, Columbus, Ohio 43210, USA

§ Department of Electrical Engineering, University of Missouri-Columbia, Columbia, Missouri 65211, USA

|| Department of Hydrology & Water Resources, University of Arizona, Tucson, Arizona 85721, USA

¶ EG&G Services, NASA Wallops Flight Facility, Wallops Island, Virginia 23337, USA

Snow & Ice Research Group, University of Nebraska, Lincoln, Nebraska 68583, USA

The response of grounded ice sheets to a changing climate critically influences possible future changes in sea level. Recent satellite surveys over southern Greenland show little overall elevation change at higher elevations, but large spatial variability^{1–3}. Using satellite studies alone, it is not possible to determine the geophysical processes responsible for the observed elevation changes and to decide if recent rates of change exceed the natural variability. Here we derive changes in ice-sheet elevation in southern Greenland, for the years 1978–88, using a physically based model of firn densification⁴ and records of annual snow accumulation reconstructed from 12 ice cores at high elevation. Our patterns of accumulation-driven elevation change agree closely with contemporaneous satellite measurements of ice-sheet elevation change, and we therefore attribute the changes observed in 1978–88 to variability in snow accumulation. Similar analyses of longer ice-core records show that in this decade the Greenland ice sheet exhibited typical variability at high elevations, well within the long-term natural variability. Our results indicate that a better understanding of ice-sheet mass changes will require long-term measurements of both surface elevation and snow accumulation.

The previously reported, 1978–88 elevation-change estimates for the southern Greenland ice sheet¹, derived from analysis of Seasat (6 July to 10 October 1978) and Geosat-Exact Repeat Mission (8 November 1986 to 7 November 1988) satellite radar altimeter data, were improved recently by adding data from the Geosat-Geodetic Mission (1 April 1985 to 30 September 1986). The updated analysis⁵ included twice as many elevation-change measurements and expanded the spatial coverage by 50%. In addition, a precise global-ocean reference network created from four years of Topex/Poseidon altimeter data was used to obtain improved corrections for altimeter radial orbit errors and measurement system biases. The updated elevation-change map includes about 90% of the area above the 2,000-m elevation contour south of 72.1° N (maximum latitude of Seasat/Geosat) and about 75% of the area west of the elevation divide between 1,700–2,000 m (Fig. 1). Effectively no coverage is available below 1,700 m because of the poor quality and limited quantity of the radar altimeter data near the ice-sheet edge⁵. The rate of elevation change (dH/dt) varies spatially from -24 to $+24$ cm yr^{-1} over distances as small as 200 km (note the strong dH/dt gradient around 66° N and the abrupt transition from positive to negative change at the elevation divide).

Between 1997 and 1999, ice cores that penetrated at least to a depth corresponding to 1978 were collected at ten locations around the southern ice sheet. All cores were analysed for a number of seasonally varying chemical species (hydrogen peroxide, calcium, ammonium, nitrate ions) using a continuous ice-core melter analysis system⁶. Two additional seasonally varying constituents, insoluble dust concentration and stable water-isotope ratios, were measured using discrete samples. In combination with field measurements of snow density, these multi-parameter glacio-chemical records were used to determine records of net annual snow accumulation at each site. Because of spatial variability in snow accumulation⁷, multiple cores were collected at four locations to identify regional-scale versus local-scale accumulation components^{8,9}. Discrete measurements of hydrogen peroxide from a core drilled in 1988 were used to develop a record of annual accumulation at Dye 3 (65.2° N, 43.9° W)¹⁰ and a 1778–1989 accumulation record was developed at Crete (71.1° N, 37.3° W) by combining published accumulation records^{11,12}.

The density of near-surface firn layers increases with depth and, in response to changing snow accumulation rates and temperature, this increase with depth also varies with time. To derive changing ice-sheet elevation from measured time series of snow accumulation, we used a model that predicts the evolution of the density–depth profile by solving coupled heat- and mass-transfer equations using a one-dimensional, lagrangian scheme, referenced to pressure coordinates⁴. The constitutive relationship for snow was calculated using rate equations for the various sintering mechanisms important in snow compaction^{13,14}. The initial depth profiles of temperature and density were obtained from steady-state calculations at the mean temperature, mean accumulation rate and mean surface density. Boundary conditions include the temperature of the ice at depth, and surface histories of temperature¹⁵, accumulation

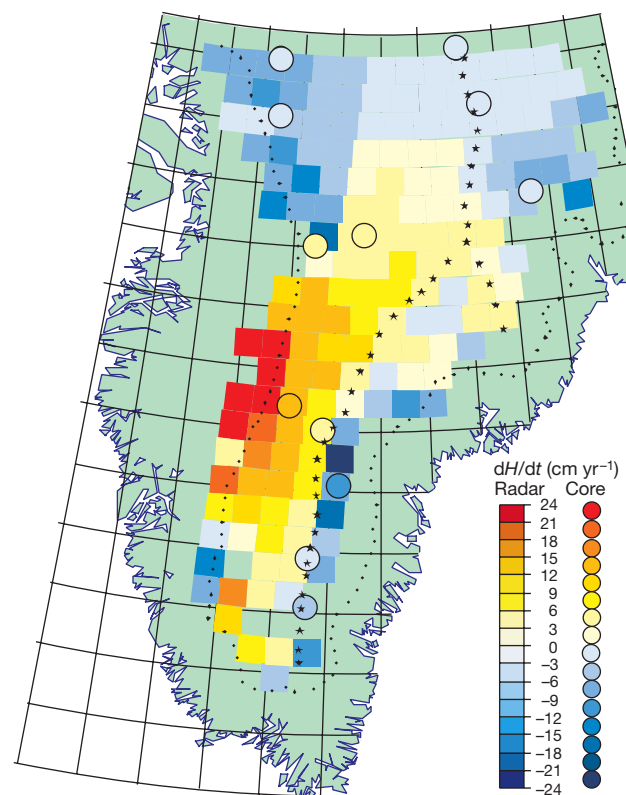


Figure 1 Accumulation-derived and radar-altimetry-derived 1978–88 ice-sheet elevation change. Shown are the approximate locations of the elevation divide (stars), the 2,000-m elevation contour (dots) and the ice-core sites (circles). The radar dH/dt values are shown in 50×50 km cells (adapted from ref. 5).

rate and density. Mean surface density was derived from the ice-core data using the zero-depth intercept of a polynomial fit to the measured density–depth profile. Densification rates for the upper two metres were also derived from the measured density profiles rather than from sintering rate equations (at shallow depths the effects of vapour transport, seasonal temperature variations and reworking by wind are important although these were not explicitly included in the sintering model). The annual temperature cycle and surface density (uppermost 1 m) in Greenland change little from year to year. Earlier calculations⁴ show that elevation changes resulting from these intra-annual variations are small compared with those resulting from inter-annual variations in snow accumulation.

The modelled change in vertical separation between the surface and the 4-MPa pressure level overburden (deep enough to lie within incompressible ice) defines the elevation change caused by variations in density⁴. The actual elevation change observed at any location is the sum of this component and the change in the local ice-equivalent-thickness (defined as the total mass of ice occupying a unit area of bedrock, divided by the density of pore-free ice). Because the divergence of the ice flow field at each core site is unknown, we cannot directly calculate the change in ice-equivalent thickness over the time period spanned by the ice core records. However, on the assumption that the ice-flow divergence is constant during this period, an estimate for the higher-frequency changes in ice-equivalent thickness can be obtained using the measured accumulation rates and added to the elevation changes predicted by the density model. We refer to the resulting time series as the ‘elevation anomaly’ at each site. These will be superimposed upon any secular trend in elevation change caused by long-term mass imbalance. Both the observed and accumulation-derived elevation changes at the twelve locations where annual accumulation data are available over the 10-year period are listed in Table 1 and shown in Fig. 1. To facilitate comparison with the radar-derived observations, 1978–88 accumulation-driven elevation change (dH/dt) is expressed in cm yr^{-1} .

We investigated the sensitivity in the calculation of accumulation-derived estimates of ice-sheet elevation change to several sources of error. To examine the sensitivity to errors in the constitutive equation for snow, strain rates in the densification model were artificially increased and decreased by a factor of ten⁴. To determine the sensitivity of our results to fluctuations in accumulation rate before the beginning of the ice-core records, we used an ensemble of six plausible ~50-year accumulation-rate histories having the same mean and inter-annual variance as the observed record in Monte-Carlo simulations. Similarly, we investigated the effects of short-scale (tens of metres) spatial variations of accumulation rate by adding noise to the measured time series and repeating these calculations for an ensemble of such records. The variance of the local noise component was obtained from closely

spaced ice cores at four of the ice-core sites^{8,9}. Note that kilometre-scale topography may also influence accumulation, although generally at multi-decadal to century, rather than annual, timescales. The 65% confidence intervals in Table 1 indicate the combined contribution from all these sources of error. Sensitivity to the accumulation history before the start of the core decreases rapidly with time under the relatively warm conditions found in Greenland⁴. Many of the core records extended to 1974 and earlier (Table 1) and for these sites, short-scale spatial variability in accumulation during the 1978–88 period provided the largest contribution to the combined uncertainties. At the six sites with shorter accumulation records, approximately half of the total uncertainty was the result of unknown accumulation before the start of the core record. If the previous accumulation were markedly different than during the core-based observation period, this would affect the results at sites with relatively short accumulation records.

The radar dH/dt values reported in Table 1 are the average of all elevation-change measurements within a circular area centred at the ice-core location. A nominal radius of 20 km was used for ten of the core locations, whereas a radius of 30 km was used at two core locations (Summit and Dye 3) where insufficient data were available within the nominal radius. Each set of data was edited by excluding elevation differences greater than two standard deviations from the primary gaussian distribution. The 65% confidence intervals include the combined contribution of random measurement error and estimates of the uncertainties in various corrections applied to the elevation data (isostatic uplift, inter-satellite bias, orbit-error corrections, and environmental corrections)^{1,5}.

Comparisons between the accumulation-derived and radar-altimeter elevation-change estimates (see Supplementary Information) indicate that the large majority of the measured change in ice-sheet elevation in southern Greenland from 1978 to 1988 is the result of temporal variability in snow accumulation. The accumulation-derived elevation change gives the same general pattern of strong thickening to the west of the elevation divide and thinning to the east. We note the excellent match in the strong radar dH/dt gradient across the ice divide from Dye 2 (66.4° N, 46.2° W) to Dye 3 (65.2° N, 43.9° W). The notable exception is in the western region along the 2,000-m contour between 69° N and 72° N where the accumulation-derived elevation change is generally smaller than the radar-based observations. A significant difference was also found in this area in a comparison between 1978–88 radar dH/dt and 1993–98 laser dH/dt results¹⁶ with no corresponding change in accumulation rate between the two time periods. Previous modelling and analysis of the Seasat and Geosat altimeter waveforms demonstrated that there was a significant temporal change in the geophysical characteristics of the near-surface snow in this particular area of the ice sheet¹⁷. Indeed, an unusually large number of melt events were found in the 69.0° N, 45.0° W core at depths corresponding to 1977 and 1978, with relatively few melt events in the early to mid-1980s. No such anomalous melt features were observed in the nearby 69.2° N, 43.0° W core where there is good

Table 1 Ice-core accumulation-derived and radar-based estimates of mean dH/dt from 1978 to 1988

Site latitude (°N), longitude (°W)	Initial year in accumulation record	1978–1988 Accumulation dH/dt (cm yr^{-1})	1978–1988 Radar* dH/dt (cm yr^{-1})
72.3, 38.0 (Summit)	1950	–1.1 (± 2.9)	–0.1 (± 1.5)
71.1, 37.3 (Crete)	1778	–0.7 (± 4.2)	–1.9 (± 1.5)
71.9, 47.5	1974	–1.9 (± 2.1)	–8.0 (± 1.5)
71.1, 47.2	1974	–1.1 (± 2.0)	–5.1 (± 2.1)
69.8, 35.0	1976	–3.8 (± 1.5)	–6.2 (± 1.5)
69.2, 43.0	1977	+5.8 (± 1.6)	+4.5 (± 1.7)
69.0, 45.0	1977	+6.9 (± 2.3)	–22.3 (± 1.7)
66.4, 46.2 (Dye 2)	1900	+12.7 (± 2.3)	+14.9 (± 1.6)
66.0, 44.5	1976	+4.8 (± 1.2)	+4.7 (± 1.6)
65.2, 43.9 (Dye 3)	1931	–10.3 (± 2.3)	–8.6 (± 1.8)
63.8, 45.0	1978	–0.4 (± 2.0)	+3.7 (± 1.6)
63.1, 44.8	1978	–3.2 (± 2.0)	–2.3 (± 1.8)

Error limits show the 65% confidence interval.

* Includes subtraction of 0.5 cm yr^{-1} uplift from long-term isostatic rebound.

Table 2 Ice core accumulation-derived estimates of the standard deviation in decadal ice-sheet elevation change (dH/dt) over the specified time ranges

Site name	Latitude (°N), longitude (°W)	Time range	Standard deviation in decadal-scale dH/dt (cm yr^{-1})
Humboldt	78.5, 56.8	1690–1994	2.4
Nasa U	73.8, 49.5	1690–1994	4.1
Summit	72.3, 38.0	1950–1998	2.0
Crete	71.1, 37.3	1778–1982	3.7
Milcent	70.3, 45.0	1177–1966	8.4
Dye2	66.4, 46.2	1900–1998	12.7
Dye3	65.2, 43.9	1931–1987	9.4

1978–88 dH/dt values for Summit, Crete, Dye 2, and Dye 3 are given in Table 1.

agreement between the radar and accumulation dH/dt estimates. These observations indicate that a large melt layer existed in 1977 and 1978 which resulted in a different radar signature at the snow–air interface than in subsequent years. This caused Geosat–Seasat dH/dt estimates to be biased towards excess thinning and/or reduced thickening in that region over the 10-year period.

To put these accumulation-derived changes in ice-sheet elevation in historical perspective, we derived estimates of dH/dt over longer time periods using previously published^{6,10,12} and new records of annual accumulation. From the annual dH/dt time series, decadal dH/dt (cm yr^{-1}) values were computed. Standard deviations in decadal dH/dt (Table 2) range from 2.0 cm yr^{-1} at the Summit location in central Greenland to 12.7 cm yr^{-1} at the southern Dye 2 site. Comparisons of the historical range of decadal dH/dt (Table 2) with the 1978–88 accumulation-derived and radar-based measurements (Table 1) indicate that recently observed changes in ice-sheet elevation above $\sim 2,000\text{-m}$ elevation are typical. Because snow accumulation at Dye 2 is highly cyclic, even the marked increase in annual accumulation from 1978 to 1988 that has led to the recent large increase in ice-sheet elevation in that region is not unusual over the last century.

These widely distributed accumulation measurements and computed elevation anomalies demonstrate that altimetry-derived estimates of ice-sheet thickening and thinning from 1978–88 over much of the southern Greenland ice sheet above $\sim 2,000\text{-m}$ elevation are consistent with elevation changes caused by temporal variability in snow accumulation. The ice-core accumulation measurements reported here do not address the significance of observed elevation change at lower elevations². Analyses of longer accumulation records indicate that the decadal-scale changes in ice-sheet elevation that occurred during 1978–88 are typical over the last few centuries and well within the natural variability of accumulation-driven elevation change. It is clear that decadal-scale variability in snow accumulation has the potential to mask longer-period changes in surface elevation associated with ice-sheet dynamics. This suggests that accurate detection of any long-term mass imbalance of the ice sheets and assessment of likely causes will require multi-decadal time series of surface elevation in conjunction with widely distributed ice-core-derived accumulation measurements collected over the time period of interest. □

Received 2 February; accepted 12 July 2000.

1. Davis, C. H., Kluever, C. A. & Haines, B. J. Elevation change of the southern Greenland Ice Sheet. *Science* **279**, 2086–2088 (1998).
2. Krabill, W. et al. Rapid thinning of parts of the southern Greenland Ice Sheet. *Science* **283**, 1522–1524 (1999).
3. Zwally, H. J., Brenner, A. C. & DiMarzio, J. P. Comment: Growth of the southern Greenland Ice Sheet. *Science* **281**, 1251 (1998).
4. Arthern, R. J. & Wingham, D. J. The natural fluctuations of firn densification and their effect on the geodetic determination of ice sheet mass balance. *Clim. Change* **40**, 605–624 (1998).
5. Davis, C. H., Kluever, C. A., Haines, B. J., Perez, C. & Yoon, Y. T. Improved elevation change measurement of the southern Greenland Ice Sheet from satellite radar altimetry. *IEEE Trans. Geosci. Remote Sens.* **38**, 1367–1378 (2000).
6. Ankin, M., Bales, R. C., Mosley-Thompson, E. & Steffen, K. Annual accumulation at two sites in northwest Greenland during recent centuries. *J. Geophys. Res.* **103**, 28775–28783 (1998).
7. Van der Veen, C. J. Interpretation of short-term ice-sheet elevation change inferred from satellite altimetry. *Clim. Change* **23**, 383–405 (1993).
8. McConnell, J. R., Mosley-Thompson, E., Bromwich, D. H., Bales, R. C. & Kyne, J. D. Interannual variations in snow accumulation on the Greenland Ice Sheet (1985–1996): New observations versus model predictions. *J. Geophys. Res.* **105**, 4039–4046 (2000).
9. Fisher, D. A., Reeh, N. & Clausen, H. B. Stratigraphic noise in time series derived from ice cores. *Ann. Glaciol.* **7**, 76–83 (1985).
10. Sigg, A. & Neftel, A. Evidence for a 50% increase in H_2O_2 over the past 200 years from a Greenland ice core. *Nature* **351**, 557–559 (1991).
11. Ankin, M., Stauffer, B., Geis, K. & Wagenbach, D. Pattern of annual snow accumulation along the west Greenland flow line: no significant change observed during recent decades. *Tellus B* **46**, 294–303 (1994).
12. Clausen, H. B., Gundestrup, N. S., Johnsen, S. J., Bindshadler, R. & Zwally, J. Glaciological investigations in the Crete area, central Greenland: A search for a new deep-drilling site. *Ann. Glaciol.* **10**, 10–15 (1988).
13. Alley, R. B. Firn densification by grain-boundary sliding: A first model. *J. Phys. (Paris)* **48**, 249–256 (1987).
14. Maeno, N. & Ebinuma, T. Pressure sintering of ice and its implication to the densification of snow at polar glaciers and ice sheets. *J. Phys. Chem.* **87**, 4103–4110 (1983).

15. Ohmura, A. New temperature distribution maps for Greenland. *Z. Gletscherkunde Glazialgeol.* **23**, 1–45 (1987).
16. Thomas, R. et al. 20-year time series of Greenland Ice-Sheet thickness change from radar and laser altimetry. *Polar Geogr.* (submitted).
17. Davis, C. H. Temporal change in the extinction coefficient of snow on the Greenland ice sheet from an analysis of Seasat and Geosat altimeter data. *IEEE Trans. Geosci. Remote Sens.* **34**, 1066–1073 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We acknowledge D. Belle-Oudry, M. Hutterli, B. Snider, E. Kline, Z. Li, P.-N. Lin, V. Zagorodnov, E. Ramsey, the Twin Otter and LC-130 aircraft crews and others for assistance in the laboratory and the field. This work was supported by grants from NASA and the US NSF.

Correspondence and requests for materials should be addressed to J.M. (email: jmconn@dri.edu).

Evidence for iron, copper and zinc complexation as multinuclear sulphide clusters in oxic rivers

Tim F. Rozan*, Michael E. Lassman†, Douglas P. Ridge† & George W. Luther III*

*College of Marine Studies, University of Delaware, 700 Pilottown Road, Lewes, Delaware 19958, USA

†Department of Chemistry and Biochemistry, Lamont duPont Laboratory, University of Delaware, Newark, Delaware 19716, USA

The availability and toxicity of trace metals in fresh water are known to be regulated by the complexation of free metal ions with dissolved organic matter^{1–3}. The potential role of inorganic sulphides in binding trace metals has been largely ignored because of the reduced persistence of sulphides in these oxic waters. However, nanomolar concentrations of copper and zinc sulphides have been observed in four rivers in Connecticut and Maryland^{4,5}. Here we report dissolved ($< 0.2 \mu\text{m}$ particle diameter) sulphide concentrations ranging up to 600 nM, with more than 90% being complexed by copper, iron and zinc. These complexes account for up to 20% of the total dissolved Fe and Zn and 45% of the total dissolved Cu. Fourier transform mass spectrometry reveals that these complexes are not simple $\text{M}(\text{HS})^+$ protonated species^{6,7} but are higher-order unprotonated clusters (M_3S_3 , M_4S_6 , M_2S_4), similar to those found in laboratory solutions^{8–10} and bio-inorganic molecules¹¹. These extended structures have high stability constants^{8,10} and are resistant to oxidation and dissociation^{10,12}, which may help control the toxicity of these and other less abundant, but more toxic, trace metals, such as silver, cadmium and mercury.

We found the total dissolved sulphide (DS_T) concentrations and metal sulphide speciation in river waters to be strongly correlated with the extent of watershed development. DS_T , defined as $\Sigma \text{Fe sulphides} + \text{Cu sulphides} + \text{Zn sulphides} + \text{other soft class B metal sulphides (Ag, Cd, Hg, Pb)} + \text{HS}^- + \text{S}_\text{x}^{2-} + \text{S}^0$, was found in concentrations ranging between 200 and 580 nM (Table 1). High DS_T ($> 200 \text{ nM}$) was found in rivers that drained urbanized areas that received inputs of treated sewage effluent. In these rivers, the metal sulphide speciation, as determined by electrochemical analysis^{4,5,13,14}, consisted of metal sulphide complexes in the following order of concentration abundance: Fe sulphides ($\Sigma \text{FeS} + \text{FeSH}^+$) $\gg$ Zn sulphides $>$ Cu sulphides $>$ (polysulphides). Polysulphides ($\text{S}_{4,5}^{2-}$), which are partly oxidized sulphide species that are reduced by Cr(II) , were only electrochemically measured¹⁵ in low

nanomolar concentrations at the two most urbanized sampling locations (Schuylkill River at Philadelphia, Pennsylvania, and Potomac River at Washington DC). In contrast, river water collected from watersheds in more rural areas (Broadkill River and Patuxent River) not affected by effluent from sewage treatment plants had low DS_T (< 50 nM) that was primarily composed ($> 88\%$) of FeS. No polysulphides were observed in these rivers. The absence of polysulphides in the majority of rivers and their presence in the two most urbanized watersheds suggests that polysulphide concentrations are only observable when substantial amounts of sewage treatment plant effluent is present in the water column.

Voltammetric analysis of effluent from two sewage treatment plants in Delaware (Table 1) revealed high concentrations of both DS_T and polysulphides. DS_T was found to exist in low micromolar concentrations, with polysulphide concentrations over 250 nM. The sulphide species concentrations in the treated sewage effluent were observed in the following order of abundance: Fe sulphides ($\Sigma FeS + FeSH^+$) $\gg$ polysulphides $\gg$ Zn sulphides $>$ Cu sulphides. The high concentrations of Fe sulphide species and polysulphides result from the reaction of sulphide with excess Fe(III) present or added during the sewage treatment process. This chemical reaction caused oxidation of sulphides to polysulphides and reduction of Fe(III) to Fe(II), which further reacts with sulphides to form both FeS and $FeSH^+$ (ref. 16).

Mass spectra obtained by Fourier transform mass spectrometry (FTMS), which were first obtained from laboratory samples of Fe, Cu and Zn sulphides, revealed different stoichiometries for these metal sulphide species, depending on the phase of the metal sulphide. Low-micromolar aqueous solutions of Cu and Zn sulphides were found to contain a wide distribution of lower-molecular-mass metal sulphides (< 400 Da) with no protons and a predominantly 1:1 metal to sulphur ratio (including MS , M_2S_2 and M_3S_3), and a small quantity of compounds with a 2:3 metal to sulphur ratio (M_4S_6). In contrast, fresh Cu and Zn sulphide precipitates and sulphur minerals (sphalerite and covellite) contained these, as well as higher-molecular-mass complexes, with

M_4S_6 and M_4S_{10} as the dominant species. As with Cu and Zn sulphides, higher-molecular-mass Fe sulphide species (that is, Fe_3S_6 and Fe_4S_8) were only associated with the fresh FeS precipitates and pyrite. In aqueous laboratory solutions, Fe_2S_4 was found to be the most abundant species. But unlike the Cu and Zn sulphides, the Fe sulphide complexes generally also contained one or two oxygen atoms and up to 5 H atoms, suggesting that H_2O molecules are still bound to the Fe(II).

In natural waters, Fe sulphides—the most abundant metal sulphide—were composed mainly of a soluble FeS cluster. Soluble FeS has a discrete voltammetric signal, and has been observed in oxic and marine pore waters^{4,17}, suggesting that FeS diffuses from the anoxic sediments to the overlying water. This natural pathway could explain the predominance of FeS in rivers that drained less-urbanized watersheds. The FTMS analysis of the river water showed the Fe sulphides to be similar to the laboratory solutions with $Fe_2S_4OH_5$ predominating (Fig. 1a.).

Cu sulphide complexes, while generally comprising less than 10% of the DS_T , accounted for 13–46% of the total dissolved Cu in the river water (this study, ref. 5; Table 1). FTMS showed Cu sulphides to be predominantly Cu_3S_3 (Fig. 1b) with minor amounts of Cu_4S_6 (Fig. 1c), matching the spectra obtained from solutions prepared in the laboratory. These empirical copper to sulphur ratios (1:1 and 2:3) are similar to those observed by voltammetry during laboratory Cu and sulphide titration experiments⁹.

Higher concentrations of Zn sulphide complexes were observed only in more highly urbanized rivers and effluent from sewage treatment plants. In river-water samples, the only Zn sulphur species positively identified by FTMS was Zn_3S_4 (a Zn_3S_3 species was also tentatively identified in one sample, but exact identification was obscured by an overlapping copper complex). The Zn_3S_4 species was found as a primary Zn sulphide complex in the FTMS spectrum from laboratory-prepared aqueous solutions with comparable abundances to Zn_3S_3 . The 1:1 Zn to sulphur ratio was also determined by voltammetry from Zn and sulphide titration experiments^{9,11}.

Table 1 Metal sulphide data

Location	pH	DS_T (μM)	Cu_T (μM)	Fe_T (μM)	Zn_T (μM)	CuS (μM)	FeS (μM)	$FeSH^+$ (μM)	ZnS (μM)	% Cu_T	% Fe_T	% Zn_T	S_2^{2-} (μM)	S_{meas}
Delaware														
Broadkill†	7.80	0.022	0.006	0.87	0.012	0.002	0.019	n.d.	n.d.	23.8	2.2	n.d.	n.d.	0.93
Brandywine§	7.36	0.39	0.022	1.8	0.15	0.007	0.35	n.d.	0.011	31.8	19.8	7.3	n.d.	0.95
Christina‡	7.18	0.48	0.019	2.6	0.21	0.005	0.42	n.d.	0.023	26.3	16.2	11	n.d.	0.94
Wil. STP	7.76	3.25	0.36	12.4	2.40	0.087	1.32	1.08	0.16	24.2	19.4	6.7	0.37	0.93
Lewes STP	7.82	1.69	0.28	4.6	1.81	0.052	0.47	0.31	0.19	18.6	17	10.5	0.5	0.90
Maryland														
Patuxent†	7.46	0.036	0.006	1.4	0.032	0.001	0.029	0.003	n.d.	12.5	2.3	n.d.	n.d.	0.91
New Jersey														
Raritan§	7.38	0.25	0.015	1.2	0.17	0.007	0.175	0.019	0.033	46.1	16.2	19.4	n.d.	0.92
Pennsylvania														
Schuylkill	6.99	0.20	0.058	1.5	0.238	0.013	0.088	0.031	0.18	22.4	7.7	7.6	0.03	0.90
Virginia														
Potomac‡	7.78	0.58	0.058	5.6	0.24	0.008	0.31	0.16	0.051	25	8.4	7.6	0.02	0.95
Connecticut*														
Hammonasset†	6.63	n.m.	0.005	0.45	0.061	0.002	n.m.	n.d.	0.001	30.3	n.m.	0.5	n.m.	n.m.
Naugatuck§	7.82	n.m.	0.05	2.2	0.336	0.008	n.m.	0.002	0.021	13	n.m.	9.7	n.m.	n.m.
Quinnipiac§	7.65	n.m.	0.032	1.8	0.169	0.013	n.m.	0.018	0.022	36.9	n.m.	4.7	n.m.	n.m.
Rhode Island¶														
Pawcatuck‡	6.98	n.m.	0.007	1.7	0.126	0.005	n.m.	0.002	0.013	53.1	n.m.	8.3	n.m.	n.m.

These data were measured in seven rivers and two sewage treatment plant (STP) effluents in the eastern United States. M_T , the dissolved ($< 0.2 \mu m$) metal fraction; %M, the percentage of M_T bound to sulphides; S_{meas} , the fraction of sulphides measured by voltammetry as compared with the total dissolved sulphide (DS_T) measured by Cr(II) reduction. In the footnote descriptions below, 'STP impact' means the water sample was collected within 16 km of a STP outfall. 'No STP impact' refers to rivers with no STP outfalls or outfalls > 50 km upstream. n.d., concentrations below 0.001 μM detection limit.

* $< 0.45 \mu m$ filtrate data taken from ref. 5 for comparison with this study (n.m., not measured).

† Low urbanization and no STP impact.

‡ Medium urbanization with no STP impact.

§ Medium urbanization with STP impact.

|| High urbanization with STP impact.

¶ Rhode Island data is also from ref. 5.

One plausible reason for the lower percentage of Zn sulphides relative to the percentage of Cu sulphides in rivers (Table 1) is the replacement of the Zn in the multinuclear sulphide cluster by Cu and other more-reactive and less-abundant soft class B metals, such as Ag, Cd, Hg and Pb. To test this hypothesis, laboratory-prepared 1 μ M solutions of Zn sulphides were titrated with equimolar solutions of (1) Cu^{2+} and (2) Cu-EDTA (representing a Cu organic complex). In the first set of titrations, voltammetry was used to measure any increases in free Zn^{2+} and decreases in Cu^{2+} , which would indicate replacement by the Cu^{2+} . In a second set of titrations, decreases in the Cu-EDTA electrochemical signal¹⁸ were also used as an indicator that Cu replaces Zn. In both cases, total replacement occurred on extremely short timescales (< 1 min). These results strongly suggest that similar replacement reactions are occurring in natural systems.

Cu sulphide clusters formed in laboratory solutions were calculated, using a mole ratio method with similar stoichiometries to those observed in natural water samples⁹, to have thermodynamic stability constants of $\log \beta = 54.7$ (Cu_3S_3) and $\log \beta = 96.4$ (Cu_4S_6). These stability constants are not corrected for the presence of Cu^+ , which results from sulphide oxidation forming a more stable sulphide complex than Cu^{2+} (refs 5, 9). The calculated thermodynamic stability constants for the clusters are significantly higher

than the conditional stability constants determined for the simpler $\text{Cu}(\text{HS})^+$ species, which has $\log \beta < 7.0$ (refs 8, 9, 13). Formation of metal sulphide clusters also leads to kinetic stability, and explains the persistence of these metal sulphide clusters in oxic waters. In laboratory studies, freshly prepared Zn sulphides were found to have a half life of more than 30 days (this work, and ref. 12), whereas in water samples the natural Zn and Cu sulphides (of unknown age) were found to persist after collection for an additional 30 and 42 days, respectively⁵.

The increased thermodynamic and kinetic stability resulting from cluster formation may be important in controlling the acute toxicity of dissolved Cu (and other reactive, soft class B metals) to planktonic and microbial organisms; such organisms take up metals across membranes, and the control of toxicity would be caused by a decrease in the pool of available metals^{19,20}. Laboratory studies have clearly shown that dissolved organic matter significantly reduces acute Cu toxicity to microorganisms by forming strong complexes²¹. In fact, Cu chelators produced by marine cyanobacteria in periods of high Cu stress were found to have predicted thermodynamic stability constants of $\log \beta = 16\text{--}38$ (ref. 18), which is significantly lower than the thermodynamic stability constant for Cu sulphide clusters. Furthermore, our experiments show that Cu even when bound to EDTA ($\log \beta = 17.94$) quickly forms metal

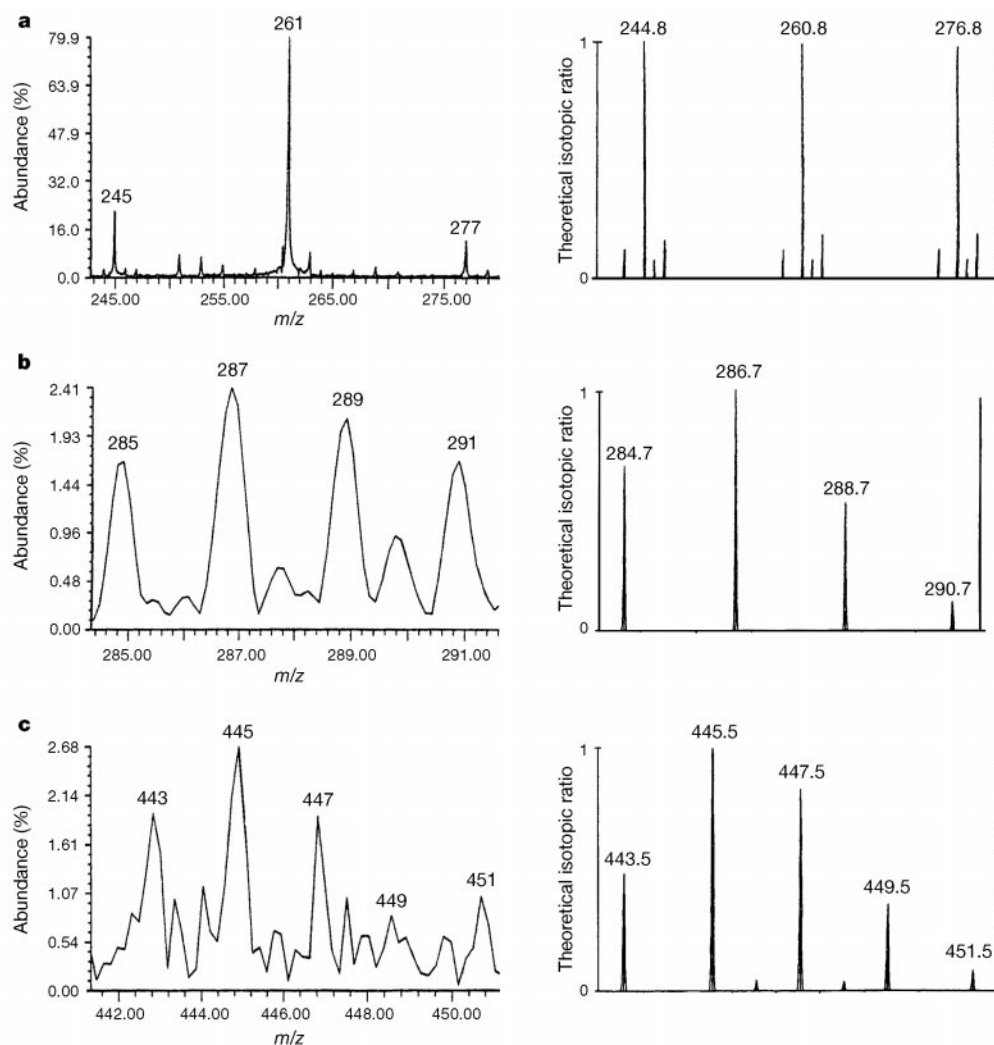


Figure 1 Representative FTMS data obtained from the Broadkill River at Milton, Delaware. **a**, $\text{Fe}_2\text{S}_4\text{H}_5$ (molecular mass 245 daltons), $\text{Fe}_2\text{S}_4\text{OH}_5$ (261 daltons) and $\text{Fe}_2\text{S}_4\text{O}_2\text{H}_5$ (277 daltons); **b**, Cu_3S_3 ; and **c**, Cu_4S_6 . The abundance (%) is relative to the most abundant anion (SO_4^{2-}) found in the sample. A graph of the theoretical isotopic ratio is given for

comparison. For natural samples, the values range from 0 to 1, with the most abundant isotopic mass given a value of 1 and the other isotopic masses plotted in proportion according to their natural abundances.

sulphide complexes, indicating that Cu sulphide clusters are more stable than Cu organic complexes. This explains why laboratory cultures of oceanic phytoplankton have been observed to increase the production of total dissolved sulphides when the concentrations of free Cu and Zn in the culture media were increased²². Although the data that we report here suggest that metal sulphide formation is a means of detoxifying trace metals for organisms, further toxicological studies are needed to quantify the roles both sulphides and 'natural' organic ligands play in controlling Cu toxicity in natural waters.

Sulphur complexation may have a dramatic effect on the acute toxicity of other, less abundant, class B metals (for example, Ag, Cd, Hg). The speciation of these trace metals should be dominated by sulphide complexation, because Zn and Fe sulphides (acid volatile sulphides) will provide a pool of sulphides available for complexation and metal replacement reactions. In laboratory experiments, we found that Ag⁺ ions quickly replaced Zn ions in metal sulphide clusters, suggesting that acid volatile sulphides are available for reaction with class B metals. Indeed, this chemistry has been observed for Cd in freshwater and marine sediments: added Cd replaced Fe in solid FeS, which resulted in a dramatic reduction in the Cd toxicity to a variety of organisms²³. □

Methods

Water samples were collected following clean protocols, and filtered in a class 100 clean room using 0.2-µm Nuclepore filters. DS_T concentrations were determined by square-wave voltammetry on a hanging-drop mercury electrode following Cr(II) reduction in acid¹⁵. Acidic Cr(II) reduction of the water samples produced H₂S, which was purged and trapped in 1 M NaOH, and then analysed by voltammetry. Acidic Cr(II) reduction was selected for its ability to dissolve both Zn and Fe sulphides and reductively dissolve pyrite and the Cu(poly)sulphides produced during Cu sulphide formation⁵⁹. Sulphide liberated with 3 M HCl was also measured, and resulted in similar sulphide concentrations to the Fe and Zn sulphide concentrations measured electrochemically. Total dissolved metals were measured using either graphite furnace atomic absorption spectrometry or inductively coupled mass spectrometry.

Metal sulphide identification was performed following the square-wave voltammetry procedure⁹ using a DLK-100 voltammetric analyser (Analytical Instrument Systems) and a 6-mm glassy carbon rotating-disk electrode (ROTEL). In summary, an acid titration was performed on the water sample to induce dissociation of the metal sulphide complexes. Free sulphide released was electrochemically measured at discrete pH values, which correspond to Cu sulphide (pH < 5), Zn sulphide (pH < 6.7) and FeSH⁺ (pH > 6.7) dissociation. To prevent double counting, the free sulphide was purged before the next acidification. FeS (ref. 13) and polysulphides (ref. 14) were measured directly by their discrete peak potentials at ambient pH.

Laser ablation FTMS was performed using a Finnegan FTMS 2000 laser desorption mass spectrometer run in negative-ion mode. The laser source was a CO₂ laser with an output at 10.6 µm wavelength. The ionization delay was set at either 1 or 5 s. FTMS analysis was performed on vacuum-dried aqueous solutions, fresh precipitates, and crushed minerals. Laboratory solutions were prepared using 1 µM aliquots of M²⁺ and HS⁻ at pH 7.5. Filtered river water samples were freeze-dried before FTMS analysis. The effects of freeze-drying appear to be minimal. Aqueous solutions were found to primarily contain lower-molecular-mass species, unlike either fresh precipitates or sulphide minerals. Additional corroborating evidence was provided by electrospray FTMS analysis of an aqueous solution of Zn sulphides, which showed Zn₃S₄ as the primary metal sulphide species²⁴.

FTMS spectra were analysed using the OPUS utilities isotope program²⁵ to calculate isotopic distributions for discrete metal sulphide stoichiometries. Isotope ratios used for species determinations were ¹H 99.98% and ²H 0.02%. Oxygen ratios were ¹⁶O 99.76%, ¹⁷O 0.04% and ¹⁸O 0.2%. Sulphur ratios were ³²S 95.02%, ³³S 0.75%, ³⁴S 4.21% and ³⁶S 0.02%. Cu ratios were ⁶³Cu 69.17% and ⁶⁵Cu 30.83%. Zn ratios were ⁶⁴Zn 48.6%, ⁶⁶Zn 27.9%, ⁶⁷Zn 4.1%, ⁶⁸Zn 18.8% and ⁷⁰Zn 0.6%. Fe ratios were ⁵⁴Fe 5.8%, ⁵⁶Fe 91.72%, ⁵⁷Fe 2.2% and ⁵⁸Fe 0.28%. Abundance (%) was related to the largest species present in the water sample, which was typically SO₄²⁻.

Received 14 February; accepted 4 July 2000.

- Breault, R., Colman, J., Aiken, G. & McKnight, D. Copper speciation and binding by organic matter in copper-contaminated streamwater. *Environ. Sci. Technol.* **30**, 3477–3486 (1996).
- Xue, H., Oestreich, A., Kistler, D. & Sigg, L. Free cupric ion concentrations and Cu complexation in selected Swiss lakes and rivers. *Aquat. Sci.* **58**, 69–87 (1996).
- Xue, H. & Sunda, W. Comparison of [Cu²⁺] measurements in lake water determined by ligand exchange and cathodic stripping voltammetry and by ion-selective electrode. *Environ. Sci. Technol.* **31**, 1902–1909 (1997).
- Roazan, T. F., Benoit, G. & Luther III, G. W. Measuring metal sulfide complexes in oxic river waters with square wave voltammetry. *Environ. Sci. Technol.* **33**, 3021–3026 (1999).
- Roazan, T. F. & Benoit, G. Geochemical factors controlling free Cu ion concentrations in river water. *Geochim. Cosmochim. Acta* **19/20**, 3311–3319 (1999).
- Dyrssen, D. Sulfide complexation in surface seawater. *Mar. Chem.* **24**, 143–153 (1988).

- Al-Farawati, R. & van den Berg, C. M. G. Metal-sulfide complexation in seawater. *Mar. Chem.* **63**, 331–352 (1999).
- Luther III, G. W., Rickard, D. T., Theberge, S. M. & Olroyd, A. Determination of metal (bi)sulfide stability constants of Mn²⁺, Fe²⁺, Co²⁺, Ni²⁺, Cu²⁺, and Zn²⁺ by voltammetric methods. *Environ. Sci. Technol.* **30**, 671–679 (1996).
- Helz, G. R., Charnock, J. M., Vaughan, D. J. & Garner, C. D. Multinuclearity of aqueous copper and zinc bisulfide complexes—an EXAFS investigation. *Geochim. Cosmochim. Acta* **57**, 15–25 (1993).
- Luther III, G. W., Theberge, S. M. & Rickard, D. T. Evidence for aqueous clusters as intermediates during zinc sulfide formation. *Geochim. Cosmochim. Acta* **19/20**, 3159–3169 (1999).
- Peters, J. W., Lanzilotta, W. N., Lemon, B. J. & Seefeldt, L. C. X-ray crystal structure of the Fe-only hydrogenase (Cpl) from *Clostridium pasteurianum* to 1.8 angstrom resolution. *Science* **282**, 1853–1858 (1998).
- Vazquez, F., Zhang, J.-Z. & Millero, F. J. Effect of metals on the rate of the oxidation of H₂S in seawater. *Geophys. Res. Lett.* **16**, 1363–1366 (1989).
- Theberge, S. M. & Luther III, G. W. Determination of electrochemical properties of a soluble aqueous FeS species present in sulfidic solutions. *Aquat. Geochem.* **3**, 1–21 (1997).
- Henneke, E., Luther III, G. W., deLange, G. J. & Hoefs, J. Sulphur speciation in anoxic hypersaline sediments from the eastern Mediterranean Sea. *Geochim. Cosmochim. Acta* **61**, 307–321 (1996).
- Roazan, T. F., Theberge, S. M. & Luther III, G. W. Quantifying elemental sulfur (S⁰), bisulfide (HS⁻) and polysulfides (S_n²⁻) using a voltammetric method. *Anal. Chim. Acta* **415**, 175–184 (2000).
- Luther III, G. W. Pyrite synthesis via polysulfide compounds. *Geochim. Cosmochim. Acta* **55**, 2839–2849 (1991).
- Rickard, D., Oldroyd, A. & Cramp, A. Voltammetric evidence for soluble FeS complexes in anoxic estuarine muds. *Estuaries* **22**, 693–701 (1999).
- Croot, P. L., Moffett, J. W. & Luther III, G. W. Polarographic determination of half-wave potentials for copper-organic complexes in seawater. *Mar. Chem.* **67**, 219–232 (1999).
- Sunda, W. G. & Huntsman, S. A. Processes regulating cellular metal accumulation and physiological effects: Phytoplankton as model systems. *Sci. Tot. Environ.* **219**, 165–181 (1998).
- Knauer, K., Behra, R. & Sigg, L. Adsorption and uptake of copper by the green alga *Scenedesmus subspicatus* (Chlorophyta). *J. Phycol.* **33**, 596–601 (1997).
- Kim, S. D., Ma, H. Z., Allen, H. E. & Cha, D. K. Influence of dissolved organic matter on the toxicity of copper to *Ceriodaphnia dubia*: Effect of complexation kinetics. *Environ. Toxicol. Chem.* **18**, 2433–2437 (1999).
- Walsh, R. S., Cutter, R. S., Dunstan, W. M., Radford-Knoery, J. & Elder, J. T. The biogeochemistry of hydrogen-sulfide-phytoplankton production in the surface ocean. *Limnol. Oceanogr.* **39**, 941–948 (1994).
- Ditoro, D. M. et al. Acid volatile sulfide predicts the acute toxicity of cadmium and nickel in sediments. *Environ. Sci. Technol.* **26**, 96–101 (1992).
- Theberge, S. M. *Investigations of Metal-sulfide Complexes and Clusters: a Multimethod Approach*. Thesis, Univ. Delaware (1999).
- OPUS utility isotope program, version 3.1. (Micromass Co., Manchester, 1993).

Acknowledgements

This work was supported by an NSF Postdoctoral award (T.F.R.), the Photographic & Imaging Manufacturers Association, Inc., and a NOAA Sea Grant (G.W.L.).

Correspondence and requests for materials should be addressed to T.F.R. (e-mail: trozan@udel.edu) or G.W.L. (e-mail: luther@udel.edu).

Collapse and recovery of marine fishes

Jeffrey A. Hutchings

Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Over-exploitation and subsequent collapse of marine fishes has focused attention on the ability of affected populations to recover to former abundance levels^{1–3} and on the degree to which their persistence is threatened by extinction^{4,5}. Although potential for recovery has been assessed indirectly¹, actual changes in population size following long-term declines have not been examined empirically. Here I show that there is very little evidence for rapid recovery from prolonged declines, in contrast to the perception that marine fishes are highly resilient to large population reductions^{6,7}. With the possible exception of herring and related species that mature early in life and are fished with highly selective equipment, my analysis of 90 stocks reveals that many gadids (for example, cod, haddock) and other non-clupeids (for example, flatfishes) have experienced little, if any, recovery as much as

15 years after 45–99% reductions in reproductive biomass. Although the effects of overfishing on single species may generally be reversible¹, the actual time required for recovery appears to be considerable. To exempt marine fishes from existing criteria used to assign extinction risk^{6,7} would be inconsistent with precautionary approaches to fisheries management and the conservation of marine biodiversity.

Worldwide overfishing has raised concerns that extraordinary collapses in abundance may significantly increase the extinction probability of targeted⁴ (for example, Atlantic cod, *Gadus morhua*) and incidentally harvested⁸ (for example, barndoor skate, *Raja laevis*) marine fishes^{5,6}. This is reflected by the work of national and international agencies responsible for assigning risk categories to potentially endangered species. Present and past fisheries for Atlantic cod, whose over-exploitation and collapse have been well documented^{2,9–11}, provide one example. On the basis of global estimates of decline, the International Union for Conservation of Nature (IUCN) listed⁴ Atlantic cod as ‘vulnerable’ in 1996, the same status applied 2 years later by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC)¹². Other marine fishes assigned to risk categories by these agencies include the Pacific sardine (*Sardinops sagax*; Clupeidae)¹², haddock (*Melanogrammus aeglefinus*; Gadidae)⁴, Bering wolffish (*Anarhichas orientalis*; Anarhichadidae)¹², and the pleuronectid flatfishes Atlantic halibut (*Hippoglossus hippoglossus*) and yellowtail flounder (*Pleuronectes ferrugineus*)⁴.

There are, however, concerns that quantitative listing criteria based on temporal trends in abundance, such as those applied by IUCN¹³ and COSEWIC¹⁴, may significantly over-estimate extinction threats to marine fishes and should be modified to account for

the great natural variability in abundance, high reproductive potential and remarkable ability to recover from population collapse perceived to be characteristic of marine fishes, relative to other taxa^{6,7}. Theoretical analyses also suggest that there is nothing intrinsic to the population dynamics of exploited marine fishes that would prevent them from increasing at low population sizes¹, although the assumption that their per capita reproductive success increases at low population levels may not be as general as previously thought³.

I made use of the most comprehensive numerical fisheries data base available, maintained by R. A. Myers, Department of Biology, Dalhousie University, Halifax, Canada, at <<http://fish.dal.ca/welcome.html>>. I recorded the largest 15-year percentage decline in mature fish biomass experienced by each stock and subsequent population sizes 5, 10 and 15 years thereafter. The 15-year interval was considered short enough to obtain a reasonably large sample of populations, and long enough to be biologically meaningful, approximating the three-generation time period specified by quantitative at-risk criteria used by species-listing organizations such as IUCN and COSEWIC.

After a decline, any increase in population size, N , could be interpreted as some sort of recovery. Graphically, recovery t years after a 15-year decline can be determined from a plot of N_{t+15}/N_0 on the ordinate against magnitude of population decline on the abscissa, that is, $1 - N_{15}/N_0$. On such a plot, populations exhibiting no recovery, that is, $N_{t+15} = N_{15}$, would fall on a straight line with

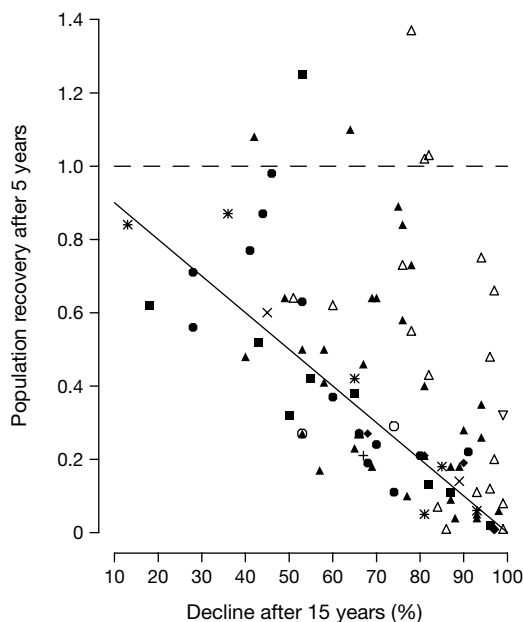


Figure 1 Bivariate association between population decline and subsequent population size for 90 marine fish stocks. The ordinate refers to the size of a population 5 years after the proportionately largest 15-year decline experienced by that population, relative to its size at the beginning of its 15-year decline. Populations that experienced some recovery are represented by points to the right of the solid line. Fully recovered stocks are represented by points on and above the dashed line. One datum—a 69% population decline of one clupeid followed by a 1.78 recovery—has been omitted for clarity. Slanted crosses, Engraulidae; upward triangles, Clupeidae; downward triangles, Osmeridae; filled triangles, Gadidae; stars, Scorpaenidae; upright crosses, Anoplopomatidae; filled diamonds, Sparidae; diamonds, Nototheniidae; filled squares, Scombridae; filled circles, Pleuronectidae; circles, Soleidae.

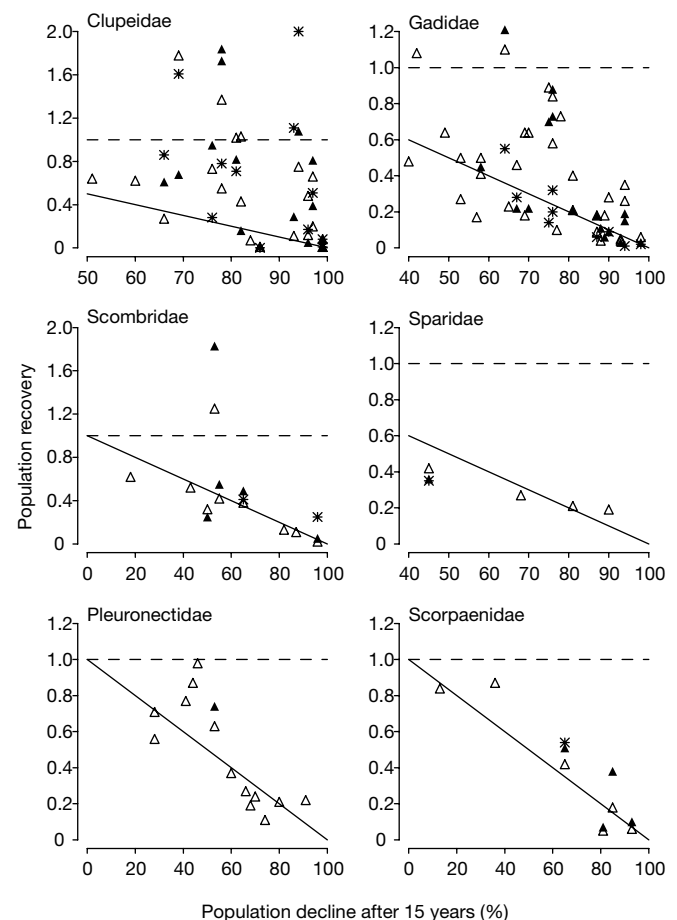


Figure 2 Population recovery within and among six families of marine fishes 5 (triangles), 10 (filled triangles) and 15 (stars) years after the greatest proportionate 15-year decline experienced by each stock. The dashed and solid lines represent the ‘full recovery’ and ‘no recovery’ lines, respectively.

slope of -1 extending from $(0,1)$ to $(1,0)$. Those continuing to decline would fall below this line, whereas those exhibiting some recovery would fall above this line. Similarly, populations falling on the line with slope of 0 extending from $(0,1)$ to $(1,1)$ would be said to have fully recovered, that is, $N_{t+15} = N_0$, whereas those above this second line would have increased relative to the size from which their population declines had begun.

Among those for which data were available, 90 marine fish stocks (representing 38 species among 11 families) experienced 15-year declines of 13 to 99%, followed by 5-year changes in population size ranging from 0.3 to 178% of the size from which the declines began. Of these 90 stocks, 37 (41%) continued to decline after the 15-year period, 46 (51%) exhibited some recovery, and 7 (8%) had fully recovered (Fig. 1). Subsequent population size was negatively correlated with magnitude of population decline (Fig. 1, Table 1; correlation coefficients will have been over-estimated slightly because of autocorrelated data).

Population sizes 5 years thereafter were significantly correlated with population declines within most of the six numerically dominant families in the analysis (representing 83 of the 90 stocks; Fig. 2, Table 1): Clupeidae (for example, herring, *Clupea harengus*; sprat, *Sprattus sprattus*), Gadidae (for example, cod, haddock), Scombridae (for example, mackerel, *Scomber* spp.; tuna, *Thunnus* spp.), Sparidae (for example, snapper, *Pagrus auratus*), Scorpaenidae (for example, redfish, *Sebastes* spp.), and Pleuronectidae (for example, plaice, *Pleuronectes* spp.; Greenland halibut, *Reinhardtius hippoglossoides*). The association was not significant within the clupeids or the scombrids, although the correlation was highly significant within the latter family when the outlier stock (eastern Pacific yellowfin tuna, *T. albacares*) was removed from the analysis. Among families, clupeids (80% of the stocks) were most likely to have experienced some level of recovery (Fig. 2).

Data 10 and 15 years subsequent to the 15-year population declines were available for 45 and 25 stocks, respectively, comprised primarily of clupeids and gadids (69 and 84%, respectively). Among all stocks, the magnitude of decline negatively influenced population size 10 years, but not 15 years, after the declines (Fig. 3, Table 1). However, when clupeid data were excluded, population decline was strongly and significantly associated with both 10- and 15-year population recovery sizes among non-clupeids, notably gadids. Indeed, 15 years after their declines, 12% of marine stocks (all clupeids) had exhibited full recovery, whereas 40% (primarily gadids, but some clupeids) had experienced no recovery at all (Fig. 3). Furthermore, when clupeids were excluded from the analyses the regression coefficients for the 5- and 15-year recovery

intervals (Table 1) were very close to the no-recovery line intercept and slope of 1 and -1 , respectively, emphasizing the limited recoveries experienced by most marine fishes.

These data suggest that, after prolonged decline, clupeids are more likely to recover to previously experienced population sizes and are more resilient than other marine fishes. Such an increased rate of recovery may be attributable to the younger age at which clupeids mature relative to gadids, scorpaenids, scombrids, sparids and pleuronectids¹⁵, and the higher intrinsic rate of increase that earlier maturity generally effects¹⁶. Higher reproductive rates may also mitigate the negative influence of environmental stochasticity on the persistence of populations at small sizes^{17,18}. In addition, being at a lower trophic level than other families considered here, clupeids may be better able to 'track' temporal and spatial fluctuations in primary and secondary productivity.

The apparent increased resilience of many clupeids relative to other stocks, however, may reflect different management responses to population declines and the different species selectivities of fishing gear. Clupeids are typically fished by deploying purse seines or mid-water trawls on schools identified visually or acoustically. The species uniformity of clupeid schools results in bycatches of incidentally harvested fishes one to two orders of magnitude lower than those of the bottom-deployed seines and trawls used to catch groundfish¹⁹. Thus, it may be comparatively easy to eliminate fishing mortality on affected clupeid stocks because of the high species selectivity of clupeid fishing technology. By contrast, the collapse of a groundfish stock rarely results in cessation of bottom-trawling in the affected region, meaning that fishing mortality on the collapsed stock can be reduced, but rarely eliminated, because of the comparatively low species selectivity of the fishing gear used to capture a broad diversity of demersal marine fishes.

The failure of many marine fish stocks to recover rapidly to former levels of abundance might arguably^{6,7} be attributed to management strategies designed to maintain populations at 50% of their virgin biomass to maximize sustainable yields²⁰. But this seems a highly improbable explanation for the present observations, given the lengthy histories of direct and indirect exploitation^{21,22} that have preceded the formal collection of data on fish numbers and biomass by fishery management agencies. In addition to the proposed influences of age at maturity, reproductive rate and fishing gear selectivity on recovery rates, it seems likely that ecosystem-level consequences of exploitation, for example, food webs altered by changes to species community structure, are also important^{23,24}.

The suggestion that marine fishes should be exempt from existing

Table 1 Linear regressions for associations between recovery and decline in marine fishes

Recovery period	Family	Regression equation	Number of stocks	P	r
5 years	Clupeidae	$y = 1.78 - 1.48x$	20	0.062	-0.42
	Gadidae	$y = 1.14 - 1.00x$	31	0.003	-0.52
	Scombridae	$y = 0.80 - 0.79x$	8	<0.001	-0.96
	Sparidae	$y = 0.65 - 0.53x$	4	0.013	-0.99
	Scorpaenidae	$y = 1.11 - 1.13x$	6	0.004	-0.95
	Pleuronectidae	$y = 1.16 - 1.19x$	13	0.0013	-0.79
	All	$y = 1.00 - 0.82x$	90	<0.0001	-0.46
	All except Clupeidae	$y = 1.04 - 0.94x$	70	<0.0001	-0.64
10 years	Clupeidae	$y = 2.74 - 2.46x$	15	0.083	-0.46
	Gadidae	$y = 1.86 - 1.87x$	16	0.007	-0.65
	All	$y = 1.41 - 1.19x$	45	0.012	-0.37
	All except Clupeidae	$y = 1.46 - 1.42x$	30	0.001	-0.56
15 years	Clupeidae	$y = 3.01 - 2.47x$	12	0.33	-0.32
	Gadidae	$y = 1.22 - 1.28x$	9	0.002	-0.88
	All	$y = 1.06 - 0.60x$	25	0.58	-0.12
	All except Clupeidae	$y = 0.91 - 0.87x$	13	0.004	-0.74

Regression equations describing mature population size (y) 5, 10 and 15 years following a 15-year population decline relative to the population size at the beginning of the decline (N_0), as a function of population decline (x , proportional decline in mature fish biomass over a period of 15 years, that is, $1 - N_{15}/N_0$). The regression equation for the Scombridae excludes the outlying eastern Pacific yellowfin tuna stock for which the 5-year recovery population size was 1.25 times the maximum during the 15-year decline (Fig. 2). Inclusion of this datum yields the following regression for the scombrids: $y = 0.99 - 0.93x$, $P = 0.078$, $r = -0.62$. Low P -values indicate rejection of the hypothesis of full recovery.

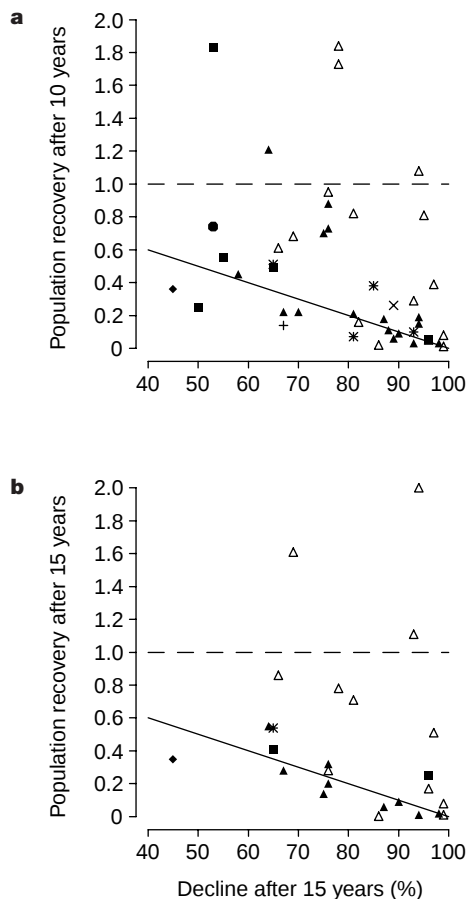


Figure 3 Population recovery of marine fishes 10 (upper panel) and 15 years (lower panel) after the greatest proportionate 15-year decline experienced by each stock. The dashed and solid lines represent the 'full recovery' and 'no recovery' lines, respectively. The families corresponding to each symbol are those given in the caption to Fig. 1. **a**, 10 years; **b**, 15 years after decline.

quantitative criteria used to assign extinction risk^{6,7} would be inconsistent with a precautionary approach to fisheries management and to the conservation of marine biodiversity. This is particularly important if, rather than providing estimates of extinction probability, the primary utility of at-risk designations lies in their reflection of the likelihood that species or populations will recover to former levels of abundance. I have found that 5–15 years after 15-year declines of 50 and 80% (the three-generation thresholds for the IUCN's 'endangered' and 'critically endangered' categories), gadid and other non-clupeid populations, on average, have increased marginally or not at all. Thus, although the effects of overfishing may indeed be generally reversible¹, the time required for population recovery in many marine fishes appears to be considerably longer than previously believed. □

Methods

Data used in the analyses presented here were available for the following families, species, and stocks/populations at (<http://fish.dal.ca/welcome.html>). Clupeidae: Atlantic menhaden, *Brevoortia tyrannus* (west Atlantic); herring, *Clupea harengus* (central British Columbia (BC), Downs, east Bering Sea, Gulf of Maine, Hokkaido, International Council for the Exploration of the Sea (ICES) VIa, Iceland (spring and summer spawners), Northwest Atlantic Fisheries Organization (NAFO) 4–5, North Sea, north and south Strait of Georgia (BC), Norway (spring spawners), Prince Rupert (BC), Queen Charlotte Island (BC), southeastern Alaska); Spanish sardine, *Sardinops sagax* (California, South Africa); and sprat, *Sprattus sprattus* (Baltic 26–28). Engraulidae: anchovy, *Engraulis encrasicolus* (Black Sea); and northern anchovy, *E. mordax* (California). Gadidae: Atlantic cod, *Gadus morhua* (Baltic (22/24 & 25–32), Greenland offshore, ICES VIa, Iceland, NAFO (2J3KL, 3NO, 3Pn4RS, 3Ps, 4TVn, 4VsW & 4X), northeast Arctic, North Sea, West Greenland); haddock, *Melanogrammus aeglefinus* (ICES VIa, Iceland, NAFO (4TVW, 4X & 5Z), NE

Arctic, North Sea); whiting, *Merlangius merlangius* (western Black Sea, North Sea); silver hake, *Merluccius bilinearis* (NAFO 5Ze, Mid-Atlantic Bight); pollock, *Pollachius virens* (ICES VI, Iceland, northeast Arctic, North Sea); and Norway pout, *Trisopterus esmarkii* (North Sea). Nototheniidae: icefish, *Notothenia rossii* (South Georgia). Scombridae: chub mackerel, *Scomber japonicus* (southern California); king mackerel, *Scomberomorus cavalla* (western Gulf of Mexico); albacore tuna, *Thunnus alalunga* (south Pacific); yellowfin tuna, *T. albacares* (east Pacific, Indian Ocean); southern bluefin tuna, *T. maccoyii* (south Pacific); bigeye tuna, *T. obesus* (east Pacific, west Atlantic); and Atlantic bluefin tuna, *T. thynnus* (west Atlantic). Sparidae: New Zealand snapper, *Pagrus auratus* (Hauraki Gulf, New Zealand SNA 8); and yellow sea bream, *Tauti tumifrons* (central China Sea, Japan). Pleuronectidae: petrale sole, *Eopseta jordani* (southern British Columbia); American plaice, *Hippoglossoides platessoides* (NAFO 3LNO); Pacific halibut, *Hippoglossus stenolepis* (north Pacific); common dab, *Limanda limanda* (Belt Sea); longhead dab, *L. propinqua* (western Kamchatka Shelf); flounder, *Platichthys flesus* (Baltic 24/25); yellowtail flounder, *Pleuronectes ferrugineus* (NAFO 5Z); plaice, *P. platessa* (Irish Sea, Kattegat, North Sea); and Greenland halibut, *Reinhardtius hippoglossoides* (east Bering Sea, northeast Arctic). Soleidae: sole, *Solea vulgaris* (ICES VIIId, North Sea). Anoplopomatidae: sablefish, *Anoplopoma fimbria* (western United States). Scorpaenidae: Pacific ocean perch, *Sebastes alutus* (Aleutian Islands, Goose Island Gully (BC), Gulf of Alaska); shortspine thornyhead, *S. alkanus* (Gulf of Alaska); and redfish, *Sebastes* spp. (Iceland, northeast Arctic). Osmeridae: caplin, *Mallotus villosus* (Barents Sea).

Received 29 February; accepted 20 June 2000.

- Myers, R. A., Barrowman, N. J., Hutchings, J. A. & Rosenberg, A. A. Population dynamics of exploited fish stocks at low population levels. *Science* **269**, 1106–1108 (1995).
- Cook, R. M., Sinclair, A. & Stefánsson, G. Potential collapse of North Sea cod stocks. *Nature* **385**, 521–522 (1997).
- Leirmann, M. & Hilborn, R. Depensation in fish stocks in a hierarchical Bayesian meta-analysis. *Can. J. Fish. Aquat. Sci.* **54**, 1976–1984 (1997).
- The International Union for Conservation of Nature (IUCN). *IUCN Red List of Threatened Animals* (IUCN, London, 1996).
- Carlton, J. T., Geller, J. B., Reaka-Kudla, M. J. & Norse, E. A. Historical extinctions in the sea. *Annu. Rev. Ecol. Syst.* **30**, 515–538 (1999).
- Hudson, E. & Mace, G. (eds) *Marine Fish and the IUCN Red List of Threatened Animals* (Zoological Society of London, London, 1996).
- Musick, J. A. Criteria to define extinction risk in marine fishes. *Fisheries* **24**, 6–12 (1999).
- Casey, J. M. & Myers, R. A. Near extinction of a large, widely distributed fish. *Science* **228**, 690–692 (1998).
- Hutchings, J. A. & Myers, R. A. What can be learned from the collapse of a renewable resource? Atlantic cod, *Gadus morhua*, of Newfoundland and Labrador. *Can. J. Fish. Aquat. Sci.* **51**, 2126–2146 (1994).
- Myers, R. A., Hutchings, J. A. & Barrowman, N. J. Hypotheses for the decline of cod in the North Atlantic. *Mar. Ecol. Prog. Ser.* **138**, 293–308 (1996).
- Sinclair, A. F. & Murawski, S. A. in *Northwest Atlantic Groundfish: Perspectives on a Fishery Collapse* (eds Boreman, J., Nakashima, B. S., Wilson, J. A. & Kendall, R. L.) 71–93 (American Fisheries Society, Bethesda, Maryland, 1997).
- Shank, C. C. The Committee on the Status of Endangered Wildlife in Canada (COSEWIC): a 21-year retrospective. *Can. Field-Nat.* **113**, 318–341 (1999).
- IUCN Species Survival Commission. *IUCN Red List Categories* (IUCN, Cambridge, 1999).
- Campbell, R. R. Rare and endangered fishes and marine mammals of Canada: COSEWIC Fish and Marine Mammal Subcommittee status reports: X. *Can. Field-Nat.* **110**, 454–461 (1996).
- Froese, R. & Pauly, D. (eds) *FishBase 99* [online] (cited 18 Feb 00) (<http://www.fishbase.org>) (1999).
- Roff, D. A. *The Evolution of Life Histories: Theory and Analysis* (Chapman & Hall, New York, 1992).
- Lande, R. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* **142**, 911–927 (1993).
- Sæther, B.-E., Engen, S., Swenson, J. E., Bakke, Ø. & Sandegren, F. Assessing the viability of Scandinavian brown bear, *Ursus arctos*, populations: the effects of uncertain parameter estimates. *Oikos* **83**, 403–416 (1998).
- Alverson, D. L., Freeberg, M. H., Murawski, S. A. & Pope, J. G. A global assessment of fisheries bycatch and discards. *FAO Fish. Tech. Pap.* **339**, 1–233 (1994).
- Hilborn, R. & Walters, C. *Quantitative Fisheries Stock Assessment*. (Chapman & Hall, New York, 1992).
- Hutchings, J. A. & Myers, R. A. in *The North Atlantic Fisheries: Successes, Failures, and Challenges* (eds Arnason, R. & Felt, L.) 38–93 (Institute of Island Studies, Charlottetown, Canada, 1995).
- Smith, T. D. *Scaling Fisheries: The Science of Measuring the Effects of Fishing, 1855–1955* (Cambridge Univ. Press, Cambridge, UK, 1994).
- Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. Fishing down marine food webs. *Science* **279**, 860–863 (1997).
- Jennings, S. & Kaiser, M. J. The effects of fishing on marine ecosystems. *Adv. Mar. Biol.* **34**, 201–352 (1998).

Acknowledgements

This work was supported by a Natural Sciences and Engineering Research Council of Canada Research Grant. I thank B. Hall, I. McLaren, J. Shepherd and D. Swain for their helpful comments on previous versions of the manuscript.

Correspondence and requests for materials should be addressed to J.A.H. (e-mail: jhutch@mscs.dal.ca).

The *Drosophila* Netrin receptor Frazzled guides axons by controlling Netrin distribution

Masaki Hiramoto^{†‡}, Yasushi Hiromi^{†‡}, Edward Giniger[§] & Yoshiki Hotta^{*‡||}

^{*} Department of Developmental Genetics, National Institute of Genetics and
[‡] Department of Genetics, Graduate University for Advanced Studies, 1111 Yata,
Mishima, Shizuoka 411-8540, Japan
[§] Fred Hutchinson Cancer Research Center, 1100 Fairview Ave, North, Seattle,
Washington 98109, USA
[†] CREST fellow, ^{||} CREST Researcher of Japan Science and Technology
Corporation, 4-1-8, Honcho, Kawaguchi 332-0012, Japan

Netrin is a secreted protein that can act as a chemotropic axon guidance cue^{1,2}. Two classes of Netrin receptor, DCC^{3–5} and UNC-5 (refs 6–9), are required for axon guidance^{3,4,6–11} and are thought to mediate Netrin signals in growth cones through their cytoplasmic domains^{12,13}. However, in the guidance of *Drosophila* photoreceptor axons, the DCC orthologue Frazzled³ is required not in the photoreceptor neurons but instead in their targets, indicating that Frazzled also has a non-cell-autonomous function¹⁴. Here we show that Frazzled can capture Netrin and 'present' it for recognition by other receptors. Moreover, Frazzled itself is actively localized within the axon through its cytoplasmic domain, and thereby rearranges Netrin protein into a spatial pattern completely different from the pattern of *Netrin* gene expression. Frazzled-dependent guidance of one pioneer neuron in the central nervous system can be accounted for solely on the basis of this ability of Frazzled to control Netrin distribution, and not by Frazzled signalling. We propose a model of patterning mechanism in which a receptor rearranges secreted ligand molecules, thereby

creating positional information for other receptors.

In vitro chemotropic responses of growth cones to Netrin indicate that graded distribution of Netrin may be important for guiding axons *in vivo*^{1,2,15–22}. A Netrin gradient could be produced by constant secretion followed by diffusion and degradation. However, we found that in the ventral nerve cord of the *Drosophila* embryo the distribution of Netrin protein cannot be explained by such a mechanism. *Drosophila* Netrin is encoded by two genes, *Netrin-A* and *Netrin-B*^{17,18}. Although Netrin messenger RNA is abundant in the midline and the ventral region of the nerve cord, Netrin-A (data not shown) and Netrin-B proteins localize in the dorsolateral region, where no Netrin mRNA is detected (Fig. 1a–d). Even when *Netrin-B* transcription was artificially restricted to midline cells, Netrin-B still accumulated in the dorsolateral region as in wild-type embryos, rather than forming a gradient centred at the midline (Fig. 1e, f; arrow). This suggests that Netrin is either transported to the dorsolateral region or is selectively captured there after secretion.

Frazzled is a good candidate for a molecule that relocates Netrin. Its accumulation is most evident on axon stalks of the commissural region³ (Fig. 1g), and its orthologue, DCC, is known to bind Netrin⁵. Moreover, the dorsolateral Netrin-positive region precisely matches Frazzled distribution (Fig. 1b, g). In the absence of Frazzled, Netrin did not accumulate dorsolaterally and Netrin-B was observed only on cell bodies that express *Netrin-B* mRNA (Fig. 1h). Moreover, when we misexpressed Frazzled in ventral unpaired median (VUM) cells, ectopic Netrin-B protein was found on their surface (Fig. 1j, k) even though these cells do not express *Netrin-B* (Fig. 1i–k). These data indicate that ectopic Frazzled can capture Netrin synthesized elsewhere, and suggest that Frazzled localizes Netrin in the dorsolateral region of ventral nerve cord. As expected, Frazzled distribution was unaltered in *Netrin-A*, *Netrin-B* double-mutant embryos (data not shown).

Frazzled itself is not found uniformly throughout the membrane, but is concentrated in specific regions of the axon (Fig. 2a, b),

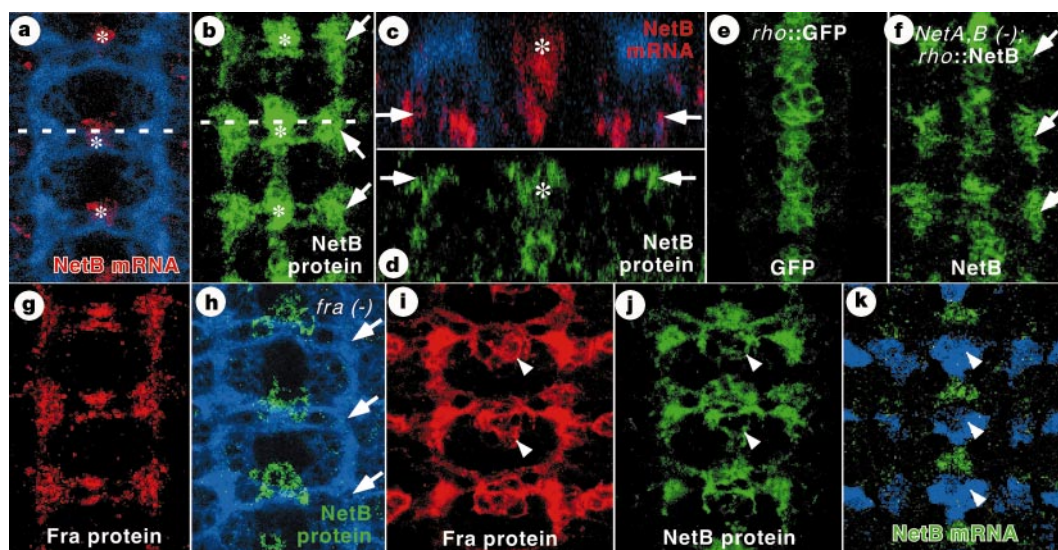


Figure 1 Frazzled relocates Netrin. Horizontal view (a, b, e–k) and cross-section (c, d) of the ventral nerve cord of stage 13 embryos. Axons were visualized using anti-HRP antibody (a, c, e, h; blue). a–d, Wild-type distribution of *Netrin-B* mRNA (a, c; red) and Netrin-B protein (b, d; green). Asterisk, midline cells. Dotted lines in a, b indicate the positions of vertical sections in c, d, respectively. *Netrin-B* mRNA is abundant ventrally (c; arrow), whereas Netrin-B protein accumulates dorsolaterally (b, d; arrows). This region corresponds to the point where commissural axons turn longitudinally. *Netrin-A* mRNA is expressed more dorsally than *Netrin-B* mRNA. Localization of Netrin-A is indistinguishable from that of Netrin-B in wild type and under experimental conditions

described in h, j and Fig. 4b (data not shown). e, f, Expression of Netrin-B in midline cells in *Netrin-A*, *Netrin-B* double-mutant embryos (*Df* (1) *NP5*; *rhomboid*–*GAL4*/+; *UAS*–*Netrin-B*/+). Projections of a series of 2- μ m confocal sections. Outlines of *rhomboid*⁺ midline cells (e) visualized with membrane-targeted GFP (GAP–GFP). g, Frazzled in wild-type embryo. Compare with distribution of Netrin B (b). h, Netrin B (green) in *frazzled* (*fra*) mutant (*fra*²/*fra*⁴). Immunoreactivity is present only near cells that express *Netrin-B* mRNA (compare with a). i–k, Ectopic Frazzled (i; red) expressed in VUM cells (arrowhead, *engrailed*–*GAL4*/+ *UAS*–*lacZ*/+ *UAS*–*fra*). Compare Netrin-B protein (j, green) and mRNA (k, green). *engrailed*–*GAL4*⁺ cells are visualized with β -galactosidase (k; blue).

indicating that its distribution may also be regulated. Localized distribution within the neuron has been observed for Roundabout (Robo), a transmembrane receptor for another guidance molecule, Slit²³, and the localization signal of Robo has been mapped to its cytoplasmic or transmembrane domain¹³. Similarly, we found that Frazzled lacking its cytoplasmic domain (Fra-ΔC) was distributed throughout the cell membrane (Fig. 2d). Furthermore, Robo-Fra, a chimera with the extracellular and transmembrane domain of Robo and the cytoplasmic domain of Frazzled, was distributed in the same way as full-length Frazzled (Fig. 2e). This shows that the cytoplasmic domain of Frazzled is necessary and sufficient for proper localization. We also expressed Fra-Robo, a chimera of the extracellular domain of Frazzled and the transmembrane and cytoplasmic domains of Robo, in *frazzled* animals. In such embryos, the Fra-Robo fusion protein failed to distribute in the wild-type Frazzled pattern, and Netrin-B was mislocalized to many of the sites of Fra-Robo accumulation (Fig. 2f). These data show that Frazzled captures Netrin with its extracellular domain, whereas Frazzled distribution is controlled by a localization signal in the cytoplasmic domain.

Next, we investigated how axons are guided by the Netrin that is captured by Frazzled. We focused on an identified pioneer neuron, dMP2, that requires *Netrin-A/Netrin-B* and *frazzled* function. dMP2 axons extend laterally and then turn posteriorly to form the initial longitudinal axon pathway^{24,25}. Precisely at the turning point, the medial edge of dorsolateral Netrin accumulation abuts the dMP2 pathway (Fig. 3d). dMP2 axons make pathfinding errors in both *Netrin-A*, *Netrin-B* double mutants and *frazzled* mutants (Fig. 3b, c, h, i), and such defects are often accompanied by severe disorganization of longitudinal tracts^{3,17,18}. These data may indicate that dMP2 axon guidance by Frazzled and Netrin is essential for the formation of the longitudinal axon pathway.

To investigate how Frazzled functions in the guidance of dMP2, we tested whether *frazzled* is required in the dMP2 neuron itself.

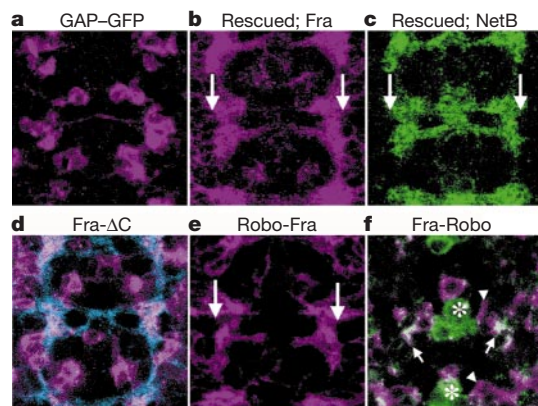


Figure 2 Functional domains of the Frazzled molecule. Frazzled and its derivatives were misexpressed using *GAL4* line G4-605. Confocal images of stage 13 embryos at the level of commissures. **a**, Outlines of *GAL4*⁺ cells visualized using GAP-GFP. **b**, **c**, Frazzled (**b**) and Netrin-B (**c**) distribution in an embryo in which full-length Frazzled expression was driven in *frazzled* background (*fra*³, G4-605/*fra*⁴; *UAS-fra*⁺). Frazzled does not distribute throughout the axon, but concentrates at specific regions (arrows, compare with **a**). Netrin-B (**c**) colocalizes with Frazzled in the lateral region, and is also present on the midline where it is independent of *frazzled* function (see Fig. 1h). **d**, **e**, Myc-tagged Frazzled derivatives expressed in wild-type background. Fra-ΔC (**d**; purple) distributes uniformly along the cell membrane, whereas Robo-Fra (**e**; purple) localizes to a specific region of the axon, as does endogenous Frazzled (blue). **f**, Expression of Fra-Robo in a *frazzled* background (*fra*², G4-605/*fra*⁴; *UAS-fra-robo*⁺). Fra-Robo (purple) is localized by the localization signal of Robo¹³. Netrin-B (green) accumulates in a subset of the Fra-Robo-positive region (arrow). Some of the highly Fra-Robo-positive regions (arrowhead) fail to accumulate Netrin-B even though they are close to the Netrin-B source at the midline (asterisk).

Contradictory to the idea that Frazzled is a Netrin sensor in dMP2 growth cones, Frazzled protein was not detected in dMP2 (Fig. 3e, f). Moreover, expressing Frazzled in dMP2 in a *frazzled* background did not rescue the defects in dMP2 axon guidance (Fig. 3k, l). In contrast, when we expressed Frazzled in many central neurons in the *frazzled* background, the defects of dMP2 axon guidance were rescued (Fig. 3m), even though Frazzled was not expressed in dMP2 (Fig. 2b). These data indicate that, for this guidance decision, Frazzled acts as a pathway marker and not as a sensor in growth cones.

The ability of Frazzled to capture Netrin raises the possibility that Frazzled guides dMP2 by capturing and presenting Netrin to dMP2.

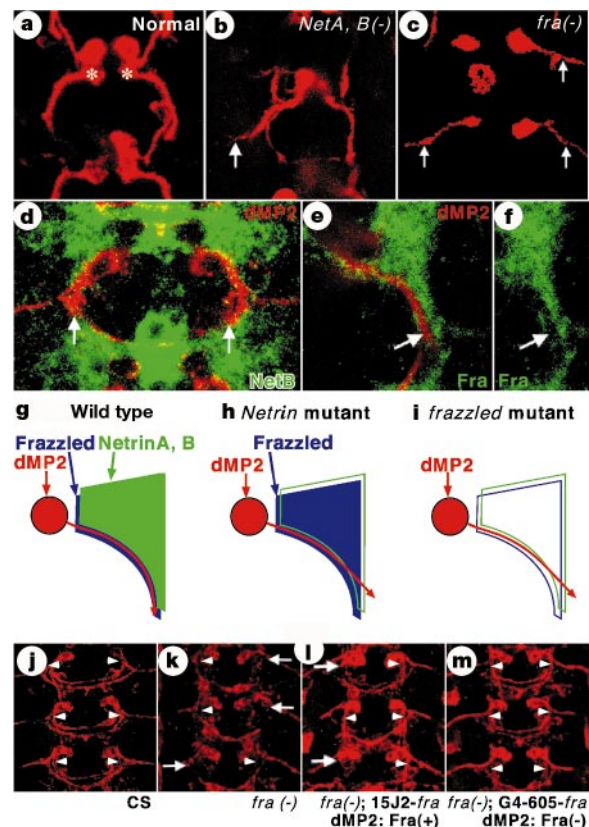


Figure 3 Axonal guidance of dMP2. dMP2 axons were visualized by driving tau-β-galactosidase expression using 15J12 *GAL4* driver (**a**–**c**; **e**), or by staining with monoclonal antibody 22C10 (**d**, **j**–**m**). **a**–**c**, dMP2 axons in wild-type embryos, (**a**) in *Netrin-A*, *Netrin-B* double mutants (**b**; *Df(1)T9B118*; 7.1%, *n* = 269) and *frazzled* mutants (**c**; *fra*²/*fra*⁴; 30.3%, *n* = 135). In these mutants dMP2 axons often extend laterally (arrows), turn anteriorly or stall (not shown). Asterisk, dMP2 cell body. **d**–**f**, In early stage 13 wild-type embryos, dMP2 growth cones (red, arrow) track along the medial surface of the Netrin-B-positive region (**d**; green). dMP2 does not express Frazzled (**e**, **f**; green), and the axon path (**e**, **f**; arrow) is devoid of Frazzled. **g**–**i**, Spatial relationships between dMP2 axon trajectory (red) and domains of Netrin (green) and Frazzled (blue). In wild-type embryos, dMP2 growth cones track the medial surface of the dorsolateral Netrin-A and Netrin-B region created by Frazzled. In *Netrin-A*, *Netrin-B* double mutants (**h**) and *frazzled* mutants (**i**), neither Netrin-A nor Netrin-B is present in the dorsolateral region and dMP2 axons make pathfinding errors. It is not clear why the *Netrin-A*, *Netrin-B* double mutant produces less penetrant phenotypes than *frazzled* mutants. We note, however, that this is also true for the commissural phenotypes of these mutants. It is possible that Frazzled localizes additional guidance molecules. **j**–**m**, dMP2 axonal defect in *frazzled* mutant is non-cell-autonomous. Arrowheads, dMP2 axons extending along the correct pathway; arrows, those making pathfinding errors in late stage 13 embryos. **j**, Wild type (CS). In *frazzled* mutant (**k**; *fra*²/*fra*⁴), many dMP2 axons make pathfinding errors (25%, *n* = 112). Restoration of Frazzled expression in dMP2 (**l**; *fra*²/*fra*⁴; *UAS-fra*/15J2) does not rescue the axonal phenotype (18.8%, *n* = 144), but expression in the dMP2 path (**m**; *fra*²/*fra*⁴; G4-605/*UAS-fra*) does (0.8%, *n* = 128).

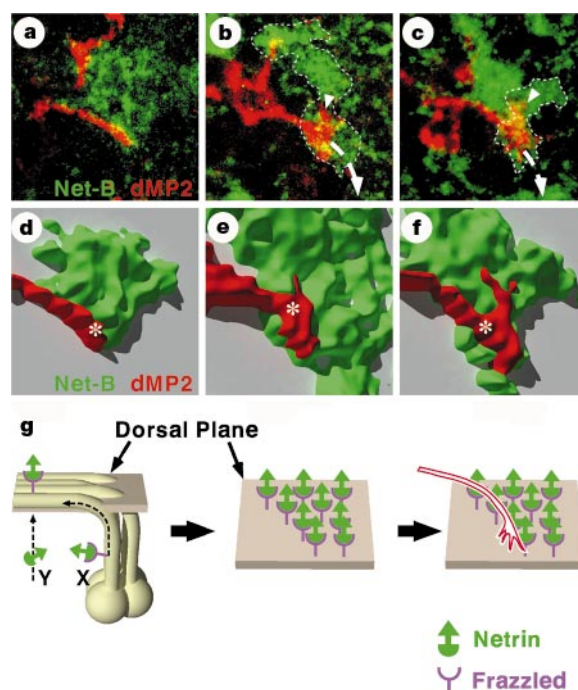


Figure 4 dMP2 growth cones respond to ectopic Netrin-B. The behaviour of the dMP2 growth cones in wild type (**a**) and upon ectopic expression of Fra-ΔC (**b**; *UAS-fra-ΔC/G4-605*) and Netrin-B (**c**; *UAS-Netrin-B/G4-605*). dMP2 growth cones (red) visualized by monoclonal antibody 22C10 in stage 12/0 embryos. Green, Netrin-B. Projections of series of 0.5-μm confocal sections. **b**, Expression of Fra-ΔC in a *frazzled* mutant background (*fra³, G4-605/fra⁴, UAS-fra-ΔC*) created an artificial Netrin-B-positive region (dotted line). dMP2 growth cone spreads abnormally (arrowhead) over the ectopic Netrin-B surface (55.8%, *n* = 52). Wild-type (**a**) and *frazzled* mutant (*fra³/fra⁴*) dMP2 growth cones never show such behaviour. Dotted arrow, normal path of the dMP2 axon (**b**, **c**). **c**, Misexpression of Netrin-B in the dMP2 axon pathway. dMP2 growth cone responded to ectopic Netrin-B (dotted contour) by spreading its growth cone anteriorly (arrowhead, 37.9%, *n* = 108). Occasional failure of dMP2 to show abnormal growth cone morphology

To test this, we expressed Fra-ΔC in a *frazzled* mutant background to create an ectopic Netrin-B positive region near the axon pathway of dMP2 without changing the pattern of *Netrin* transcription. In such embryos, dMP2 growth cones spread abnormally over this surface of artificial Netrin accumulation (Fig. 4b). We also directly misexpressed Netrin-B in cell bodies located near the dMP2 axon pathway. Again, dMP2 growth cones responded to the ectopic Netrin-B-positive region (Fig. 4c). This strongly suggests that the response of dMP2 to the ectopic Frazzled extracellular domain is due to a response to the Netrin bound to the domain.

An implication of these data is that dMP2 uses a Netrin receptor other than Frazzled to respond to Netrin. Redirection of dMP2 growth cones to ectopic Netrin indicates that Netrin is perceived as an attractive cue to dMP2. As the *Drosophila* genome does not contain any other genes with significant homology to DCC, we expect that the Netrin receptor expressed in dMP2 is structurally different from the DCC class of Netrin receptors.

Our data indicate that Frazzled captures and rearranges Netrin, and presents it to other growth cones (Fig. 4d). The capture/relocation mechanism we have uncovered can create a precise Netrin distribution even in regions that are quite distant from the source of Netrin protein. Just as Frazzled presents Netrin to the dMP2 axon at its lateral turning point, the vertebrate Frazzled orthologue DCC also captures Netrins, and is localized to the point where the commissural axons turn longitudinally⁵. Presentation of Netrin may thus be a general feature of DCC proteins. How Netrin reaches its final location is not yet clear. As Netrin-B does not localize to all Fra-Robo positive regions even when they are close to

under these experimental conditions is due to segmental variability in the timing of Netrin-B produced by the GAL4/UAS system. When the expression of ectopic Netrin-B occurs before the arrival of the dMP2 growth cone, dMP2 always responds (**b**, **c**). **d–f**, Three-dimensional views of samples shown in **a–c**, respectively, reconstructed from a series of confocal sections for 5-μm depth. dMP2 growth cones track the surface of Netrin-B-positive regions. Asterisks, corresponding positions in each sample. **g**, Capture/relocation model of Netrin and Frazzled. Plane indicates dorsal surface of ventral nerve cord, where commissures and longitudinal connectives lie. Left, Netrin is relocated to the dorsal region. This probably involves transport along axons with Frazzled (X) and diffusion (Y). Middle, Frazzled rearranges Netrin in a specific pattern. Right, dMP2 growth cone uses rearranged Netrin as a guidance cue.

a source of Netrin-B (Fig. 2f), relocation of Netrin is likely to involve transport along axons rather than diffusion alone (Fig. 4d). Perhaps active relocation of receptors such as Frazzled or Robo²³ may be used to transport ligands to the final target area, where they are interpreted by other receptors. In addition to neuronal axons, extended cellular processes, such as the cytonemes of *Drosophila* imaginal discs and vertebrate limb buds²⁶, have been implicated in other patterning systems. It will be interesting to see whether such systems also use capture/relocation mechanisms to generate precise spatial patterns away from the source of the diffusible morphogen.

Methods

Constructs and strains

The Fra-ΔC construct (Fig. 4) was made as follows. A *NotI* linker was inserted into Frazzled complementary DNA³ at codon 1,097 within the cytoplasmic region. This plasmid was linearized by *NotI*, and a myc epitope (Invitrogen) and a stop codon were added, generating the sequence ¹⁰⁹⁷LAAAEQKLISEEDLNGAAsop. The resulting Fra-ΔC fragment was cloned into pUAST²⁸ for transformation into flies. Frazzled and Robo derivatives¹³ used for the localization experiment (Fig. 2) all contain the transmembrane domain of Robo, and carry 6-myc epitopes at their carboxy termini. The following strains have been described^{3,17,25,27,28}: *fra³, fra⁴, UAS-fra, Df(1)NP5, Df(1)T9B118, UAS-Netrin-B, GAL4 line 15J2, UAS-tau-lacZ, UAS-lacZ*. GAL4 line G4-605, *engrailed GAL4* and *UAS-GAP-GFP* strains were gifts from G. M. Technau, A. Brand and A. Chiba, respectively. The G4-605 line expresses GAL4 in many central neurons, but not in dMP2.

Immunofluorescence staining

Embryos were fixed and processed for indirect immunofluorescence microscopy as described²⁹. Confocal images were taken on a Biorad MRC-1024 imaging head, mounted on a Zeiss Axioplan II microscope. Images were processed using Bio-Rad LaserSharp 2.1A,

Adobe Photoshop 4.0, Adobe Streamline 3.0 and Strata Studio Pro 2.5.3. We used the following primary antibodies: rabbit anti-Frazzled³ (1:500 dilution); goat anti-Netrin-B¹⁸ (1:500); rabbit anti- β -galactosidase (1:2,000, Cappel), goat anti-HRP (1:2,000, Cappel), monoclonal antibody 22C10 (ref. 30; 1:20, from Developmental Studies Hybridoma Bank (DSHB)), rabbit anti-GFP (1:100, Clontech); monoclonal antibody anti-Myc 9E10 (1:10, developed by J. M. Bishop, obtained from DSHB). Mutant embryos were identified using balancer chromosomes carrying β -galactosidase genes, or by staining with anti-Frazzled antibody (Fig. 2f, 4b) or anti-Netrin-B antibody (Fig. 3b). The rescue of the dMP2 phenotype by Frazzled expression in and outside dMP2 (Fig. 3j–m) was judged using the following criteria. To show the presence of the mutant phenotype (Fig. 3k, l), we scored the dMP2 pathway as abnormal if the 22C10⁺ fascicle was discontinuous at the segmental boundary. To demonstrate phenotypic rescue when Frazzled was expressed outside dMP2 (Fig. 3m), we scored as phenotypically normal hemisegments that had both of the following properties: (1) 22C10⁺ fascicle was present at the segmental boundary; and (2) an axon with dMP2-specific morphology was present after it has passed the segmental boundary. These criteria were chosen because there are several 22C10⁺ axons within the dMP2 fascicle at the segmental boundary. Our criteria underestimate both the dMP2 defect in Fig. 3k, l and the rescue of the phenotype in Fig. 3m.

Received 2 February; accepted 22 May 2000.

1. Serafini, T. *et al.* The netrins define a family of axon outgrowth-promoting proteins homologous to *C. elegans* UNC-6. *Cell* **78**, 409–424 (1994).
2. de la Torre, J. R. *et al.* Turning of retinal growth cones in a netrin-1 gradient mediated by the netrin receptor DCC. *Neuron* **6**, 1211–1224 (1997).
3. Kolodziej, P. A. *et al.* *frazzled* encodes a *Drosophila* member of the DCC immunoglobulin subfamily and is required for CNS and motor axon guidance. *Cell* **87**, 197–204 (1996).
4. Chan, S. S. *et al.* UNC-40, a *C. elegans* homolog of DCC (Deleted in Colorectal Cancer), is required in motile cells responding to UNC-6 netrin cues. *Cell* **87**, 187–195 (1996).
5. Keino-Masu, K. *et al.* Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175–185 (1996).
6. Leung-Hagstjorn, C. *et al.* UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in *C. elegans*. *Cell* **71**, 289–299 (1992).
7. Hamelin, M., Zhou, Y., Su, M. W., Scott, I. M. & Culotti, J. G. Expression of the UNC-5 guidance receptor in the touch neurons of *C. elegans* steers their axons dorsally. *Nature* **364**, 327–330 (1993).
8. Ackerman, S. L. *et al.* The mouse rostral cerebellar malformation gene encodes an UNC-5-like protein. *Nature* **386**, 838–842 (1997).
9. Leonardo, E. D. *et al.* Vertebrate homologues of *C. elegans* UNC-5 are candidate netrin receptors. *Nature* **386**, 833–838 (1997).
10. Fazeli, A. *et al.* Phenotype of mice lacking functional Deleted in colorectal cancer (Dcc) gene. *Nature* **386**, 796–804 (1997).
11. Przyborski, S. A., Knowles, B. B. & Ackerman, S. L. Embryonic phenotype of *Unc5h3* mutant mice suggests chemorepulsion during the formation of the rostral cerebellar boundary. *Development* **125**, 41–50 (1998).
12. Hong, K. *et al.* A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. *Cell* **97**, 927–941 (1999).
13. Bashaw, G. J. & Goodman, C. S. Chimeric axon guidance receptors: the cytoplasmic domains of slit and netrin receptors specify attraction versus repulsion. *Cell* **97**, 917–926 (1999).
14. Gong, Q., Rangarajan R., Seeger, M. & Gaul, U. The Netrin receptor Frazzled is required in the target for establishment of retinal projections in the *Drosophila* visual system. *Development* **126**, 1451–1456 (1999).
15. Serafini, T. *et al.* Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001–1014 (1996).
16. Kennedy, T. E., Serafini, T., de la Torre, J. R. & Tessier-Lavigne, M. Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell* **78**, 425–435 (1994).
17. Mitchell, K. J. *et al.* Genetic analysis of Netrin genes in *Drosophila*: Netrins guide CNS commissural axons and peripheral motor axons. *Neuron* **17**, 203–215 (1996).
18. Harris, R., Sabatelli, L. M., & Seeger, M. A. Guidance cues at the *Drosophila* CNS midline: identification and characterization of two *Drosophila* Netrin/UNC-6 homologs. *Neuron* **17**, 217–228 (1996).
19. Hedgecock, E. M., Culotti, J. G. & Hall, D. H. The unc-5, unc-6, and unc-40 genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*. *Neuron* **4**, 61–85 (1990).
20. Wadsworth, W. G., Bhatt, H. & Hedgecock, E. M. Neuroglia and pioneer neurons express UNC-6 to provide global and local netrin cues for guiding migrations in *C. elegans*. *Neuron* **16**, 35–46 (1996).
21. Ishii, N., Wadsworth, W. G., Stern, B. D., Culotti, J. G. & Hedgecock, E. M. UNC-6, a laminin-related protein, guides cell and pioneer axon migrations in *C. elegans*. *Neuron* **9**, 873–881 (1992).
22. Colamarino, S. A. & Tessier-Lavigne, M. The axonal chemoattractant netrin-1 is also a chemorepellent for trochlear motor axons. *Cell* **81**, 621–629 (1995).
23. Brose, K. *et al.* Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell* **96**, 795–806 (1999).
24. Jacobs, J. R. & Goodman, C. S. Embryonic development of axon pathways in the *Drosophila* CNS. II. Behavior of pioneer growth cones. *J. Neurosci.* **9**, 2412–2422 (1989).
25. Hidalgo, A. & Brand, A. H. Targeted neuronal ablation: the role of pioneer neurons in guidance and fasciculation in the CNS of *Drosophila*. *Development* **124**, 3253–3262 (1997).
26. Ramirez-Weber, F. A. & Kornberg, T. B. Cytonemes: cellular processes that project to the principal signaling center in *Drosophila* imaginal discs. *Cell* **97**, 599–607 (1999).
27. Winberg, M. L., Mitchell, K. J. & Goodman, C. S. Genetic analysis of the mechanisms controlling target selection: complementary and combinatorial functions of netrins, semaphorins, and IgCAMs. *Cell* **93**, 581–591 (1998).
28. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
29. Ashburner, M. *Drosophila, a Laboratory Manual*. (CSHL, Cold Spring Harbor, 1989).

30. Fujita, S. C., Zipursky, S. L., Benzer, S., Ferrus, A. & Shotwell, S. L. Monoclonal antibodies against the *Drosophila* nervous system. *Proc. Natl Acad. Sci. USA* **79**, 7929–7933 (1982).

Acknowledgements

We thank A. Chiba, C. S. Goodman, A. Hidalgo, P. Kolodziej, M. Seeger and G. Technau for fly strains; P. Kolodziej and M. Seeger for DNA clones and antibodies; T. Hosoya, Y. Umehara, M. Okabe and all members of the Hotta laboratory for helpful discussions; and Y. Fujioka, M. Seki, M. Sakai and C. Asaka for technical assistance. This work was funded by CREST (Y. Hotta and Y. Hiromi), the Ministry of Education, Science, Sports, and Culture of Japan, and Research for the Future Program of JSPS (Y. Hiromi) and NIH (E.G.).

Correspondence and requests for materials should be addressed to Y.H. (e-mail: yhotta@lab.nig.ac.jp).

Intracellular calcium dependence of transmitter release rates at a fast central synapse

Ralf Schneggenburger & Erwin Neher

Abteilung Membranbiophysik, Max-Planck-Institut für biophysikalische Chemie, Am Fassberg 11, D-37077 Göttingen, Germany

Calcium-triggered fusion of synaptic vesicles and neurotransmitter release are fundamental signalling steps in the central nervous system. It is generally assumed that fast transmitter release is triggered by elevations in intracellular calcium concentration ($[Ca^{2+}]_i$) to at least 100 μ M near the sites of vesicle fusion^{1–5}. For synapses in the central nervous system, however, there are no experimental estimates of this local $[Ca^{2+}]_i$ signal. Here we show, by using calcium ion uncaging in the large synaptic terminals of the calyx of Held, that step-like elevations to only 10 μ M $[Ca^{2+}]_i$ induce fast transmitter release, which depletes around 80% of a pool of available vesicles in less than 3 ms. Kinetic analysis of transmitter release rates after $[Ca^{2+}]_i$ steps revealed the rate constants for calcium binding and vesicle fusion. These show that transient (around 0.5 ms) local elevations of $[Ca^{2+}]_i$ to peak values as low as 25 μ M can account for transmitter release during single presynaptic action potentials. The calcium sensors for vesicle fusion are far from saturation at normal release probability. This non-saturation, and the high intracellular calcium cooperativity in triggering vesicle fusion, make fast synaptic transmission very sensitive to modulation by changes in local $[Ca^{2+}]_i$.

The rapidly decaying, local $[Ca^{2+}]_i$ signal that triggers vesicle fusion in presynaptic terminals cannot readily be resolved with current imaging techniques. To circumvent this problem, we applied Ca^{2+} uncaging in large presynaptic terminals, the calyces of Held^{6,7}, located in the brainstem medial nucleus of the trapezoid body. During Ca^{2+} uncaging, cytosolic $[Ca^{2+}]_i$ is elevated in a spatially homogenous manner⁸. Therefore, fluorescent Ca^{2+} indicators can be used to measure directly the biologically relevant $[Ca^{2+}]_i$ acting at the sensor for vesicle fusion^{9,10}.

Calyces were loaded through whole-cell patch pipettes with mixtures of the photolysable Ca^{2+} chelator DM-nitrophen (1 mM) and the low-affinity fluorescent Ca^{2+} indicator fura-2 FF (Fig. 1a). The transmitter release activity was monitored by recording excitatory postsynaptic currents (EPSCs) from principal neurons with a second patch pipette (Fig. 1a). Cyclothiazide (CTZ; 100 μ M) was present during most recordings to suppress the desensitization^{11,12} of AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)-type glutamate receptors, which mediate fast EPSCs in these synapses.

We used flashes of varying intensities to study the magnitude of the postsynaptic response as a function of presynaptic $[Ca^{2+}]_i$. Dim flashes that elevated $[Ca^{2+}]_i$ to about $2\ \mu\text{M}$ (Fig. 1b, blue trace) evoked small postsynaptic currents in which individual quantal events could be identified (Fig. 1c). Stronger flashes, which elevated the presynaptic $[Ca^{2+}]_i$ to roughly two- or threefold higher values (Fig. 1b, red and black traces), caused much larger EPSCs, in which individual quantal events were no longer detectable (Fig. 1d). Stable flash responses could be observed several times in a given cell pair. This is shown in Fig. 1d, in which the black and the grey trace, respectively, represent the response to the first and the seventh flash, which evoked similar presynaptic $[Ca^{2+}]_i$ changes and similar EPSCs. These large EPSCs reflect the release activity of several hundred active zones that operate in parallel at calyx of Held synapses^{13,14}.

To quantify transmitter release rates, we deconvolved the multi-quantal EPSC responses with the quantal response waveform^{15,16}. This was necessary because peak EPSC amplitudes are not a valid measure of transmitter release, unless release is highly synchronized. We estimated the quantal response waveforms by recording spontaneous miniature EPSCs (mEPSCs) under the same conditions as flash-evoked EPSCs. Average mEPSC traces (Fig. 1f) had amplitudes of $33 \pm 9.3\ \text{pA}$, and their decay was biphasic, with fast and slow time constants of 1.9 ± 0.4 and $13.4 \pm 2.6\ \text{ms}$, respectively ($n = 5$ cells). The deconvolution of flash-evoked, compound EPSCs therefore assumed such quantal response waveforms with double-exponential decays (see Methods). Deconvolution revealed peak release rates of about 10 and 60 quanta per ms (see Fig. 1e) for the flash-evoked, multi-quantal EPSCs shown in Fig. 1d.

In experiments with a higher DM-nitrophen concentration

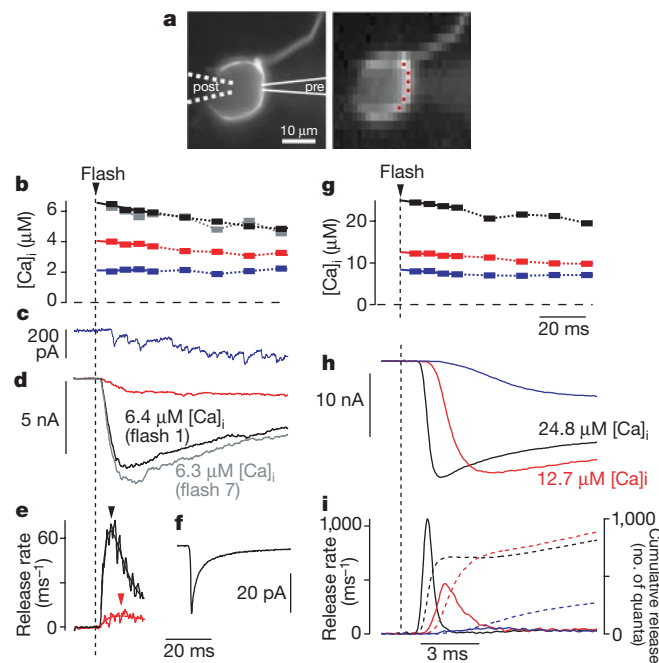


Figure 1 Presynaptic $[Ca^{2+}]$ steps and postsynaptic detection of transmitter release. **a**, Fluorescent images at 380 nm excitation of a calyx of Held nerve terminal filled through a whole-cell patch pipette with Ca^{2+} indicator, fura-2 FF and photolysable Ca^{2+} chelator, DM-nitrophen (1 mM). Red dots indicate binned pixels used for the analysis of cellular fluorescence changes. **b**, Elevations of presynaptic $[Ca^{2+}]_i$ obtained by light flashes with different transmission intensities. **c**, **d**, EPSCs in response to the presynaptic $[Ca^{2+}]_i$ elevations shown in **b**. Colours identify corresponding $[Ca^{2+}]_i$, EPSC and release rate traces in **b**–**e**. **e**, Transmitter release rates obtained from deconvolution of two EPSCs shown in **d**. **f**, An averaged mEPSC trace from 436 individual traces. **g**–**i**, Presynaptic $[Ca^{2+}]_i$ elevations (**g**), EPSCs (**h**) and transmitter release rates (**i**) for a different cell pair. In **i**, cumulative release rates (dashed lines) were obtained by integrating the corresponding release rate traces.

(2 mM), step-like elevations of up to $25\ \mu\text{M}$ $[Ca^{2+}]_i$ were produced (see Fig. 1g), which evoked larger EPSCs with faster rise times. EPSCs reached maximal amplitudes at $[Ca^{2+}]_i \approx 12\ \mu\text{M}$. Increasing $[Ca^{2+}]_i$ further caused a speeding of the rising phase, but did not alter the amplitude of EPSCs (compare black and red traces in Fig. 1h). The corresponding integrated release rate traces (Fig. 1i, dashed lines) show that at $[Ca^{2+}]_i > 12\ \mu\text{M}$, cumulative release was similar within around 3 ms after the onset of release, indicating that a pool of readily releasable vesicles^{14,17} may have been depleted by such $[Ca^{2+}]_i$ steps. The late ($>3\ \text{ms}$) slow rise in the integrated release rate traces may be due to inaccuracies in the release rate estimates during the decaying phase of EPSCs (Fig. 1h, i) or to a genuine slow component of transmitter release. A small effect of postsynaptic receptor saturation on the peak EPSC amplitudes of the largest responses cannot be excluded. This, however, should not affect our estimate of peak release rates, as these occur at the point of the steepest rise of EPSCs, before the EPSCs reach their peak amplitudes (Fig. 1h, i).

We next tested whether Ca^{2+} uncaging and the $[Ca^{2+}]_i$ signal created by the opening of voltage-gated Ca^{2+} channels induce vesicle fusion from a common pool of readily releasable vesicles. We studied the inhibition of depolarization-induced EPSCs by preceding flash-evoked transmitter release (Fig. 2)¹⁴. In the example shown in Fig. 2, a depolarization-evoked Ca^{2+} current induced a control EPSC of 25.8 nA (Fig. 2b, black trace), which was reduced to 18.1 nA (Fig. 2b, blue trace) and 4.4 nA (Fig. 2b, red trace) when the depolarizing pulses were preceded by flashes of varying intensities. This inhibition was not accompanied by a reduction in presynaptic Ca^{2+} currents (I_{Ca} was $97 \pm 5\%$ of control after flashes; $n = 9$). Half-maximal and 80% inhibition occurred with $[Ca^{2+}]_i$ steps of around

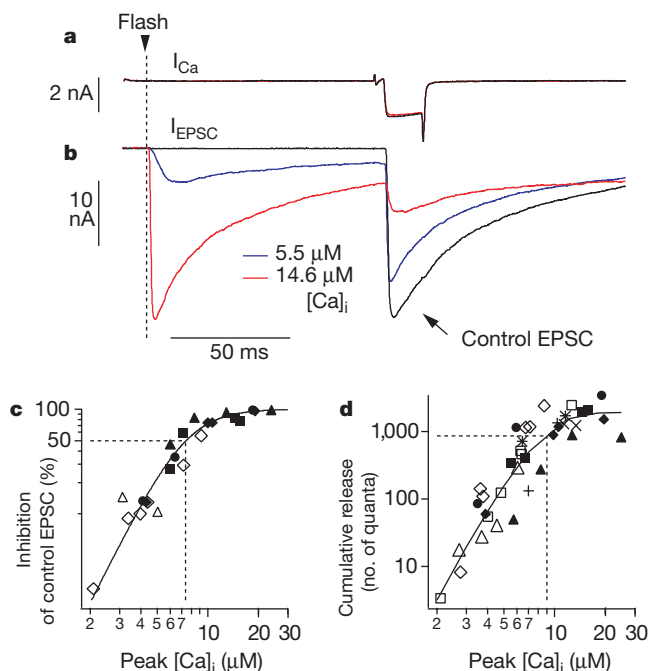


Figure 2 $[Ca^{2+}]_i$ elevations induced by flashes and presynaptic depolarizations release vesicles from a common pool. **a**, Presynaptic Ca^{2+} currents (I_{Ca}) under control conditions (black trace) and after a flash (red trace). Ca^{2+} currents were evoked by a 4-μs depolarization to +80 mV, followed by a 16-ms period at 0 mV. **b**, EPSCs corresponding to the traces shown in **a**. The blue and red traces were obtained after flashes that elevated the presynaptic $[Ca^{2+}]_i$ to the indicated levels. **c**, Inhibition of depolarization-evoked EPSCs as a function of flash-induced presynaptic $[Ca^{2+}]_i$. Half-maximal inhibition occurred at $\sim 7\ \mu\text{M}$ $[Ca^{2+}]_i$, as indicated by a Hill-function fit of the data (solid line). **d**, Cumulative transmitter release, analysed for $n = 10$ cells by integrating the release rates in a 10-ms interval (Fig. 1). The predictions of the model shown in Fig. 3e are superimposed (solid line).

7 and 10 μM , respectively (Fig. 2c). Figure 2d shows the $[\text{Ca}^{2+}]_i$ dependence of cumulative transmitter release, obtained by integrating the release rates over the first 10 ms after flashes (see Fig. 1i). The inhibition of depolarization-induced EPSCs (Fig. 2c) and the cumulative transmitter release (Fig. 2d) showed almost identical $[\text{Ca}^{2+}]_i$ dependencies. This indicates that both types of Ca^{2+} stimulus draw upon the same pool of rapidly releasable (<10 ms) vesicles. Cumulative release reached a plateau of $1,785 \pm 870$ released quanta with $[\text{Ca}^{2+}]_i$ steps above 12 μM ($n = 8$ flashes; see Fig. 2d). This number is larger than a previous pool size estimate based on analysing EPSC charges in the absence of CTZ¹⁴, perhaps owing to a contribution of AMPA-receptor desensitization¹² in the previous study. Nevertheless, the pool size estimates showed considerable scatter between cells (range 800–3,500; see Fig. 2d).

Figure 3a shows the $[\text{Ca}^{2+}]_i$ dependence of transmitter release rates, measured from 47 flash responses in 13 cell pairs. In the range 2–8 μM $[\text{Ca}^{2+}]_i$, peak release rates increased as a power function of $[\text{Ca}^{2+}]_i$, with an exponent of 4.2, as indicated by linear regression analysis in the double-logarithmic plot of Fig. 3a. This exponent is similar to previous values obtained by varying the extracellular Ca^{2+} concentration at neuromuscular junction synapses¹⁸ or at central synapses, including the calyx of Held^{13,14}. At values above 10 μM , release rates increased with a more shallow dependency on $[\text{Ca}^{2+}]_i$. In the range 8–13 μM $[\text{Ca}^{2+}]_i$, the average release rate was $349 \pm 209 \text{ ms}^{-1}$ ($n = 9$ flashes), a value similar to the peak release rates measured during action-potential-evoked EPSCs (see below; Fig. 4). Nevertheless, the corresponding average $[\text{Ca}^{2+}]_i$ ($10.7 \pm 1.5 \mu\text{M}$) is not necessarily equal to the peak $[\text{Ca}^{2+}]_i$ reached at the Ca^{2+} sensors for vesicle fusion during a single presynaptic

action potential. This is because the local $[\text{Ca}^{2+}]_i$ signal during an action potential probably decays very rapidly, which might curtail transmitter release before Ca^{2+} binding to the sensor for vesicle fusion reaches equilibrium.

To learn about the rate-limiting steps involved in Ca^{2+} -dependent vesicle fusion at the calyx of Held synapse, we fitted the flash-photolysis data with a minimal kinetic model, incorporating several Ca^{2+} -binding steps, followed by vesicle fusion from the fully Ca^{2+} -bound state^{4,10} (Fig. 3e). As it has been reported¹⁹ that uncaging of Ca^{2+} by flash photolysis can, under certain conditions, lead to brief (~ 1 –2 ms) spike-like overshoots of $[\text{Ca}^{2+}]_i$ that would have been undetected by Ca^{2+} imaging, we first modelled the expected $[\text{Ca}^{2+}]_i$ time course after flash photolysis (see Methods). The resulting $[\text{Ca}^{2+}]_i$ time course (Fig. 3d) did not show a significant spike. This is attributable to the almost full loading of DM-nitrophen with Ca^{2+} , and to rapid Ca^{2+} buffering by ATP and by endogenous Ca^{2+} buffers. The calculated $[\text{Ca}^{2+}]_i$ time course was then used to drive the Ca^{2+} -binding scheme shown in Fig. 3e. To predict the high initial slope of 4.2 in the double-logarithmic plot of release rates versus $[\text{Ca}^{2+}]_i$ (Fig. 3a), it was necessary to include five Ca^{2+} -binding steps, as well as a cooperativity factor⁴ (b in the scheme of Fig. 3e).

The analysis of synaptic delays (Fig. 3b) and times to peak release (Fig. 3c) gave important constraints for setting the Ca^{2+} -binding and -unbinding rate constants. When these rate constants were increased, a good fit of the peak release rates (Fig. 3a) was still obtained, but the delays and times to peak release (Fig. 3b, c) were then significantly underestimated at low $[\text{Ca}^{2+}]_i$ (data not shown). The set of parameters found by fitting the data in Fig. 3 also accurately predicted the decrease in cumulative transmitter release with submaximal $[\text{Ca}^{2+}]_i$ elevations (Fig. 2d). Note, however, that the value of the vesicle fusion rate γ is only a lower limit estimate because the finite rising speed of mEPSCs, as well as possible postsynaptic receptor saturation, might contribute to the apparent levelling-off of peak release rates at $[\text{Ca}^{2+}]_i > 10 \mu\text{M}$.

Knowledge of the kinetic parameters of Ca^{2+} binding and vesicle fusion (Fig. 3) should allow us to infer the time course and amplitude of the local $[\text{Ca}^{2+}]_i$ signal during transmitter release

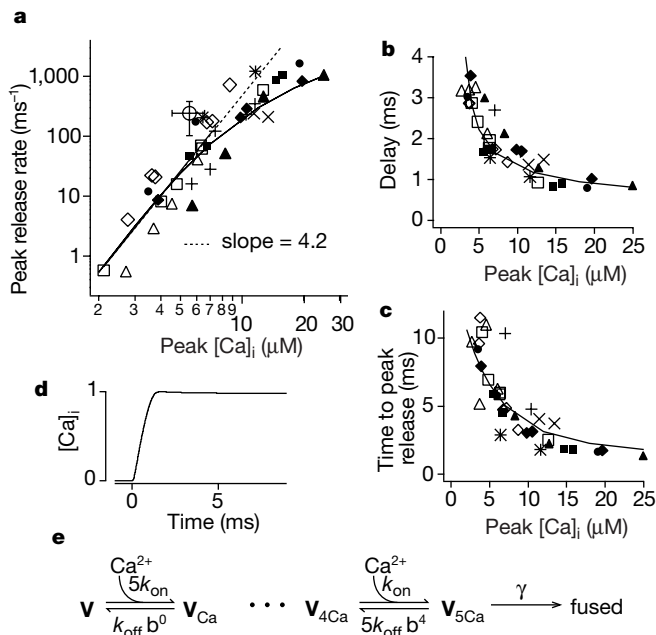


Figure 3 Intracellular Ca^{2+} concentration dependence of transmitter release rates at the calyx of Held synapse. **a**, Peak release rates as a function of presynaptic $[\text{Ca}^{2+}]_i$ after flash photolysis. Data from $n = 10$ cells with either 1 mM (open symbols) or 2 mM DM-nitrophen (filled symbols). The open round symbol represents average results ($n = 3$ cells) with NP-EGTA as photolysable Ca^{2+} chelator (see Methods). Data from experiments in the absence of CTZ (star-like symbols; $n = 2$) is also shown. The dashed line indicates a slope of 4.2. **b**, Delay between the time of flash trigger ($t = 0$), and occurrence of first transmitter release. **c**, Time to reach peak release rates (see Fig. 1e, arrowheads). In **a**–**c**, predictions of the kinetic model (**e**) are shown by solid lines. **d**, Simulation of the $[\text{Ca}^{2+}]_i$ time course after flash photolysis. **e**, Scheme of a kinetic model used to fit the data. The following parameters were found: k_{on} , $9 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$; k_{off} , $9,500 \text{ s}^{-1}$; γ , $6,000 \text{ s}^{-1}$; $b = 0.25$.

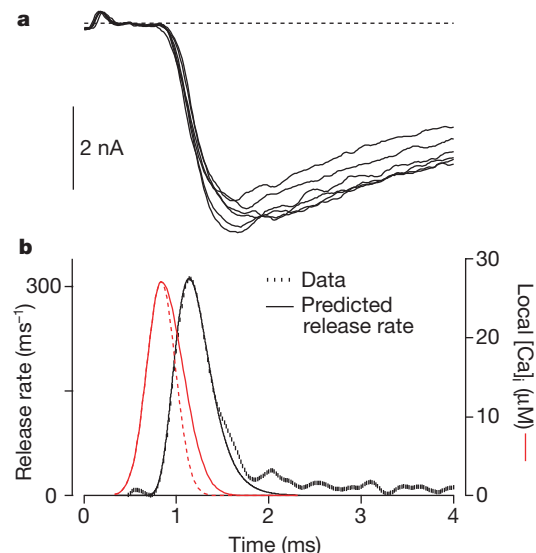


Figure 4 The predicted local $[\text{Ca}^{2+}]_i$ signal at an average Ca^{2+} sensor for vesicle fusion during single presynaptic action potentials. **a**, Six consecutive EPSCs evoked by fibre stimulation (0.05 Hz) in the presence of 2 mM Ca^{2+} , 1 mM Mg^{2+} and 100 μM CTZ. **b**, Average transmitter release rates obtained by deconvolution of the EPSCs in **a**. Red line, inferred local $[\text{Ca}^{2+}]_i$ signal that predicts a release rate (black line) similar to that observed. Dashed red curve, Gaussian with half-width of 0.36 ms, an idealized representation of the Ca^{2+} current waveform¹³ during a single action potential at the calyx of Held.

triggered by presynaptic action potentials. To do so, we first measured transmitter release rates during EPSCs evoked by afferent fibre stimulation. Six consecutive EPSCs, obtained in the presence of 100 μM CTZ, are shown in Fig. 4a on an expanded time scale. The addition of CTZ to the bath solution increased peak EPSC amplitudes by $17.7 \pm 11.1\%$ ($n = 4$), an effect that is probably caused by the more effective peak summation of mEPSCs with slowed decay times in the presence of CTZ (but see refs 16 and 20). The EPSCs were deconvolved with quantal response waveforms, which were obtained by sampling spontaneous mEPSCs between single fibre stimulations for each cell. The average peak release rate was $335 \pm 111 \text{ ms}^{-1}$ ($n = 4$ cells). The duration of transmitter release at half-maximal amplitude was $0.44 \pm 0.07 \text{ ms}$ (Fig. 4b), in good agreement with a previous estimate obtained from delay histograms¹³.

We next set out to calculate the local $[\text{Ca}^{2+}]_i$ signal that an average vesicle might experience during a presynaptic action potential. As a starting point we assumed this signal to be proportional to the Ca^{2+} current measured during action potentials at the calyx of Held¹³. This implies that the local $[\text{Ca}^{2+}]_i$ transient is produced by overlap of $[\text{Ca}^{2+}]_i$ domains created by several neighbouring Ca^{2+} channels within a short distance. We then performed model calculations with the kinetic parameters obtained from the flash data (Fig. 3), and varied the decay and the amplitude of the local $[\text{Ca}^{2+}]_i$ waveform to obtain adequate predictions of transmitter release during action potentials. In four cells with average EPSC amplitudes of $5.88 \pm 4 \text{ nA}$, we found $[\text{Ca}^{2+}]_i$ transients with peak amplitudes of $27.8 \pm 1.4 \mu\text{M}$ and half-widths of $0.46 \pm 0.05 \text{ ms}$. The width at half-amplitude of the inferred local $[\text{Ca}^{2+}]_i$ signal exceeded that of presynaptic Ca^{2+} currents during action potentials¹³ by only $\sim 100 \mu\text{s}$ (Fig. 4b, compare solid and dashed red lines), indicating that fast processes such as diffusion or binding to endogenous Ca^{2+} buffers^{3,5} must be extremely effective in terminating the local $[\text{Ca}^{2+}]_i$ signal after Ca^{2+} channels close.

We have shown that the fast phase of transmitter release at a central nervous system (CNS) synaptic terminal occurs at $[\text{Ca}^{2+}]_i$ values that are significantly lower than the range of 75–300 μM predicted by theoretical studies^{3,5}, by experiments on invertebrate synapses^{1,2,19} or by flash photolysis in retinal bipolar cell terminals⁴. In ref. 4, secretion measured by membrane capacitance was absent below 10–20 μM $[\text{Ca}^{2+}]_i$ (but see ref. 21), a Ca^{2+} concentration at which we observe strong transmitter release from calyces, at rates equalling those obtained with single action potentials (Fig. 3a). It is possible that ribbon-type synapses found in sensory systems have vesicle fusion mechanisms with a lower Ca^{2+} sensitivity than those at the calyx of Held synapse with its conventional active zones. It is usually assumed that fast synaptic transmission is driven by rapid local $[\text{Ca}^{2+}]_i$ transients to values above 100 μM , which might be found in the immediate vicinity ($< 10 \text{ nm}$) of an open Ca^{2+} channel^{3,5}. Our data show that fast vesicle fusion can occur at much lower peak $[\text{Ca}^{2+}]_i$ during single action potentials ($\sim 25 \mu\text{M}$; Fig. 4b), indicating that tight ($< 10 \text{ nm}$) molecular coupling of vesicles and Ca^{2+} channels at this synapse is not necessary for fast transmitter release. This conclusion is supported by the finding that presynaptic injection of EGTA, a Ca^{2+} buffer with a slow binding rate, reduces EPSC amplitudes at this¹³ and at neocortical synapses²², again implying that a significant contribution to the local $[\text{Ca}^{2+}]_i$ signal originates from remote channels. The possibility that the colocalization of vesicles and Ca^{2+} channels becomes more tight with maturation of synapses can, however, not be excluded at present. In any case, the local $[\text{Ca}^{2+}]_i$ at the sensors for vesicle fusion is about fifty times higher than the spatially averaged $[\text{Ca}^{2+}]_i$ attained with single action potentials ($\sim 0.5 \mu\text{M}$)²³, confirming the significance of spatially localized $[\text{Ca}^{2+}]_i$ signals for synaptic transmission^{1–5}.

Action potentials released 178 ± 86 quanta ($n = 4$ cells) within 2 ms (Fig. 4), corresponding to roughly 10% of the pool of readily releasable vesicles as determined in Fig. 2. This rather small released

fraction, which is at the lower limit of previous estimates at this synapse^{14,17}, is a typical feature of CNS synapses²⁴. Sustained $[\text{Ca}^{2+}]_i$ elevations to 25 μM (Fig. 1g, h, i; black traces) lead to transmitter release rates several times larger than those observed with single presynaptic action potentials (Fig. 4b), indicating that most Ca^{2+} sensors do not reach the fully occupied state during single presynaptic action potentials. This conclusion agrees with the finding at this synapse¹⁴ that increasing the extracellular Ca^{2+} concentration leads to marked potentiation of EPSC amplitudes. The non-saturation of the Ca^{2+} sensors during an action potential and the high intracellular Ca^{2+} cooperativity of vesicle fusion (Fig. 3a) render small changes in amplitude or time course of the local $[\text{Ca}^{2+}]_i$ signal very effective in modulating the amount of transmitter output at a synapse. Such changes in the local $[\text{Ca}^{2+}]_i$ signal are probably involved in presynaptic inhibition of transmitter release²⁵ as well as in Ca^{2+} - and activity-dependent forms of synaptic plasticity²⁶.

Note added in proof: A similar relationship between transmitter release and $[\text{Ca}^{2+}]_i$ has recently been reported using Ca^{2+} uncaging by laser photolysis³¹. □

Methods

Tissue preparation, electrophysiological recordings and solutions

We prepared transverse brainstem slices 200 μm thick from 8–10-day-old Wistar rats^{6,7,14}. Whole-cell voltage-clamp recordings were done with a double patch-clamp amplifier (EPC-9/2, HEKA elektronik) from presynaptic calyces and postsynaptic principal neurons. The holding potential was -80 mV , and pre- and postsynaptic series resistances were compensated up to 80%. For action potential stimulation of calyces, only the postsynaptic cell was recorded (Fig. 4), and afferent fibres were stimulated with a bipolar electrode. The extracellular solution was a standard Ringer solution¹⁴ with 2 mM CaCl_2 and 1 mM MgCl_2 , to which 10 mM tetraethylammonium (TEA), 0.5 μM tetrodotoxin, 50 μM D-2-amino-5-phosphonovaleric acid and 100 μM CTZ were added during paired recordings to suppress undesired current components and desensitization of AMPA receptors^{11,12}. Postsynaptic pipette (intracellular) solutions were as described¹⁴. Presynaptic pipette solutions were made as 2 $\times$ stocks containing (final concentrations in mM) 130 Cs-glucuronate, 20 TEA-Cl, 20 HEPES, 5 Na_2ATP and 0.3 Na_2GTP , to which fura-2 FF (Teflabs; 0.1 mM); DM-nitrophen (Calbiochem; 1 or 2 mM), CaCl_2 (0.85 or 1.7 mM) and MgCl_2 (0.5 or 0.8 mM) were added. For control experiments at 1 mM free intracellular Mg^{2+} (Fig. 3a), 0.1 mM fura-2FF, 5 mM NP-EGTA (Molecular Probes), 4 mM CaCl_2 , 3.3 mM MgCl_2 , 2 mM Na_2ATP and 0.3 mM Na_2GTP were added. Experiments were done at room temperature ($21\text{--}24^\circ\text{C}$). Results are reported as mean $\pm$ s.d.

Ca^{2+} uncaging and imaging

We used single light pulses ($\sim 1.5 \text{ ms}$) from a flash lamp (Rapp Optoelektronik) for photolysis. $[\text{Ca}^{2+}]_i$ imaging was resumed 4 ms after the flash trigger (Fig. 1b), by exciting fura-2 FF at 350 nm and 380 nm with a monochromator (TILL-photonics). Pixel binning (8×15 ; see Fig. 1a) was used to allow short exposure times (5 ms), and free $[\text{Ca}^{2+}]_i$ was calculated from background-corrected fluorescence ratios of a series of pixels in the calyx region (Fig. 1a). The dissociation constant ($10 \pm 2 \mu\text{M}$; $n = 5$) of the fura-2FF/ Ca^{2+} complex was first determined *in vitro*, by using mixtures of CaCl_2 and Ca^{2+} buffers (DM-nitrophen plus DPTA or HEGTA). We obtained the calibration constants at nominally 0 $[\text{Ca}^{2+}]$ (R_{min} ; 10 mM EGTA, 0 CaCl_2) and at 10 μM $[\text{Ca}^{2+}]$ ($R_{10\mu\text{M}}$) by loading calyces with calibration solutions ($n = 4$ each). For $R_{10\mu\text{M}}$ the pipette solution contained 10 μM $[\text{Ca}^{2+}]_i$ buffered by isosmolar DPTA (80 mM), and the extracellular solution was a Na^+ -free Ringer with 10 μM $[\text{Ca}^{2+}]$. $R_{10\mu\text{M}}$ was taken at the end of a 200-s recording period at a holding potential of 0 mV. The ratio at a saturating $[\text{Ca}^{2+}]$ (R_{max} ; 10 mM CaCl_2) was taken from *in vitro* measurements. The calibration constants were corrected¹⁰ for changes in DM-nitrophen fluorescence²⁷ after flashes.

Deconvolution analysis

EPSC current traces were low-pass filtered at 6 kHz and digitized at 20 kHz. Residual series resistance ($1.5\text{--}3 \text{ M}\Omega$) after partial electronic compensation was compensated off-line²⁸ before display and analysis. This led to increases in current amplitudes of up to 50%, and to small decreases in the rise times of fast EPSCs (Fig. 4). We calculated transmitter release rates from the EPSC and its time derivative according to the equation given in ref. 15, extended by a term for the slowly decaying mEPSC component (E.N. & T. Sakaba, manuscript in preparation). The mEPSC amplitude, the fast and slow decay time constants and the relative contribution of the slow component were set to the average values found from fits of mEPSCs (see text). This deconvolution analysis provides release rates averaged over time windows equal to the time-to-peak of mEPSCs. It assumes that mEPSCs add up linearly; this assumption might be violated in synapses with strong effects of glutamate spillover¹². However, a detailed analysis (E.N. & T. Sakaba, manuscript in preparation) shows that the effects of build-up of residual glutamate should be small within short times ($< 10 \text{ ms}$) after release. Delays (Fig. 3b) were expressed as the time between the triggering of the flash and a level crossing of five released quanta in integrated release rate traces. We analysed times to peak release from double-exponential fits to release rate traces (Fig. 1e). Analyses were done with IgorPro (Wavemetrics).

Kinetic simulations

The $[Ca^{2+}]_i$ time course with flashes (Fig. 3d) was simulated from the measured time course of the flash by numerical integration of a system of differential equations for the implicated Ca^{2+} and Mg^{2+} buffers. We assumed the presence of an endogenous Ca^{2+} buffer with binding capacity (κ) of 40, as determined experimentally at calyces of Held²³. The binding rate constants of DM-nitrophen, ATP and endogenous Ca^{2+} buffer were as in ref. 29. A different set of binding rate constants for the endogenous Ca^{2+} buffer³⁰ at constant κ of 40 produced similar results (data not shown). The $[Ca^{2+}]_i$ time course of Fig. 3d was used to drive the kinetic scheme for Ca^{2+} binding and vesicle fusion (Fig. 3e), which was solved numerically. Release rates were obtained by differentiating the accumulation of vesicles in the fused state, assuming a pool of 1,800 vesicles (Fig. 2d).

Received 5 May; accepted 21 July 2000.

- Adler, E. M., Augustine, G. J., Duffy, S. N. & Charlton, M. P. Alien intracellular calcium chelators attenuate neurotransmitter release at the squid giant synapse. *J. Neurosci.* **11**, 1496–1507 (1991).
- Llinás, R., Sugimori, M. & Silver, R. B. Microdomains of high calcium concentration in a presynaptic terminal. *Science* **256**, 677–679 (1992).
- Yamada, W. M. & Zucker, R. S. Time course of transmitter release calculated from simulations of a calcium diffusion model. *Biophys. J.* **61**, 671–682 (1992).
- Heidelberger, R., Heinemann, C., Neher, E. & Matthews, G. Calcium dependence of the rate of exocytosis in a synaptic terminal. *Nature* **371**, 513–515 (1994).
- Roberts, W. M. Localization of calcium signals by a mobile calcium buffer in frog saccular hair cells. *J. Neurosci.* **14**, 3246–3262 (1994).
- Forsythe, I. D. Direct patch recording from identified presynaptic terminals mediating glutamatergic EPSCs in the rat CNS, *in vitro*. *J. Physiol.* **479**, 381–387 (1994).
- Borst, J. G. G., Helmchen, F. & Sakmann, B. Pre- and postsynaptic whole-cell recordings in the medial nucleus of the trapezoid body of the rat. *J. Physiol.* **489**, 825–840 (1995).
- Naraghi, M., Müller, T. H. & Neher, E. Two-dimensional determination of the cellular Ca^{2+} binding in bovine chromaffin cells. *Biophys. J.* **75**, 1635–1647 (1998).
- Thomas, P., Wong, J. G. & Almers, W. Millisecond studies of secretion in single rat pituitary cells stimulated by flash photolysis of caged Ca^{2+} . *EMBO J.* **12**, 303–306 (1993).
- Heinemann, C., Chow, R. H., Neher, E. & Zucker, R. S. Kinetics of the secretory response in bovine chromaffin cells following flash photolysis of caged Ca^{2+} . *Biophys. J.* **67**, 2546–2557 (1994).
- Yamada, K. A. & Tang, C.-M. Benzothiazides inhibit rapid glutamate receptor desensitization and enhance glutamatergic synaptic currents. *J. Neurosci.* **13**, 3904–3915 (1993).
- Trussell, L. O., Zhang, S. & Raman, I. M. Desensitization of AMPA receptors upon multiquantal neurotransmitter release. *Neuron* **10**, 1185–1196 (1993).
- Borst, J. G. G. & Sakmann, B. Calcium influx and transmitter release in a fast CNS synapse. *Nature* **383**, 431–434 (1996).
- Schneggenburger, R., Meyer, A. C. & Neher, E. Released fraction and total size of a pool of immediately available transmitter quanta at a calyx synapse. *Neuron* **23**, 399–409 (1999).
- van der Kloot, W. Estimating the timing of quantal releases during end-plate currents at the frog neuromuscular junction. *J. Physiol.* **402**, 595–603 (1988).
- Diamond, J. S. & Jahr, C. E. Asynchronous release of synaptic vesicles determines the time course of the AMPA receptor-mediated EPSC. *Neuron* **15**, 1097–1107 (1995).
- Wu, L. -G. & Borst, J. G. G. The reduced release probability of releasable vesicles during recovery from short-term synaptic depression. *Neuron* **23**, 821–832 (1999).
- Dodge, F. A. & Rahamimoff, R. Co-operative action of calcium ions in transmitter release at the neuromuscular junction. *J. Physiol.* **193**, 419–432 (1967).
- Landó, L. & Zucker, R. S. Ca^{2+} cooperativity in neurosecretion measured using photolabile Ca^{2+} chelators. *J. Neurophysiol.* **72**, 825–830 (1994).
- Barnes-Davies, M. & Forsythe, I. D. Pre- and postsynaptic glutamate receptors at a giant excitatory synapse in rat auditory brainstem slices. *J. Physiol.* **488**, 387–406 (1995).
- Lagnado, L., Gomis, A. & Job, C. Continuous vesicle cycling in the synaptic terminal of retinal bipolar cells. *Neuron* **17**, 957–967 (1996).
- Ohana, O. & Sakmann, B. Transmitter release modulation in nerve terminals of rat neocortical pyramidal cells by intracellular calcium buffers. *J. Physiol.* **513**, 135–148 (1998).
- Helmchen, F., Borst, J. G. G. & Sakmann, B. Calcium dynamics associated with a single action potential in a CNS presynaptic terminal. *Biophys. J.* **72**, 1458–1471 (1997).
- Goda, Y. & Südhof, T. Calcium regulation of neurotransmitter release: reliably unreliable? *Curr. Opin. Cell Biol.* **9**, 513–518 (1997).
- Wu, L. -G. & Saggau, P. Presynaptic inhibition of elicited neurotransmitter release. *Trends Neurosci.* **20**, 204–212 (1997).
- Zucker, R. S. Calcium- and activity-dependent synaptic plasticity. *Curr. Opin. Neurobiol.* **9**, 305–313 (1999).
- Zucker, R. S. Effects of photolabile calcium chelators on fluorescent calcium indicators. *Cell Calcium* **13**, 29–40 (1992).
- Traynelis, S. F. Software-based correction of single compartment series resistance errors. *J. Neurosci. Methods* **86**, 25–34 (1998).
- Xu, T., Naraghi, M., Kang, H. & Neher, E. Kinetic studies of Ca^{2+} binding and Ca^{2+} clearance in the cytosol of adrenal chromaffin cells. *Biophys. J.* **73**, 532–545 (1997).
- Xu-Friedman, M. A. & Regehr, W. G. Presynaptic strontium dynamics and synaptic transmission. *Biophys. J.* **76**, 2029–2042 (1999).
- Bollmann, J. H., Sakmann, B. & Borst, J. G. G. Calcium sensitivity of glutamate release in a calyx-type terminal. *Science* (in the press).

Acknowledgements

We thank T. Sakaba for discussions, and J. Rettig, M. Casado, P. Ascher and R. Fernández-Chacon for critical comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to R.S. (e-mail: rschneg@gwdg.de).

SLAM (CDw150) is a cellular receptor for measles virus

Hironobu Tatsuo*, Nobuyuki Ono*, Kotaro Tanaka & Yusuke Yanagi

Department of Virology, Faculty of Medicine, Kyushu University, Fukuoka 812-8582, Japan

* These authors contributed equally to this work

Measles virus continues to be a major killer of children, claiming roughly one million lives a year¹. Measles virus infection causes profound immunosuppression, which makes measles patients susceptible to secondary infections accounting for high morbidity and mortality². The Edmonston strain of measles virus, and vaccine strains derived from it, use as a cellular receptor human CD46 (refs 3, 4), which is expressed on all nucleated cells; however, most clinical isolates of measles virus cannot use CD46 as a receptor⁵. Here we show that human SLAM (signalling lymphocyte-activation molecule; also known as CDw150), a recently discovered membrane glycoprotein expressed on some T and B cells⁶, is a cellular receptor for measles virus, including the Edmonston strain. Transfection with a human SLAM complementary DNA enables non-susceptible cell lines to bind measles virus, support measles virus replication and develop cytopathic effects. The distribution of SLAM on various cell lines is consistent with their susceptibility to clinical isolates of measles virus. The identification of SLAM as a receptor for measles virus opens the way to a better understanding of the pathogenesis of measles virus infection, especially the immunosuppression induced by measles virus.

A marmoset B-cell line transformed with Epstein–Barr virus (EBV), B95-8, and its adherent subline B95a are highly sensitive to measles virus (MV) present in clinical specimens⁷; and strains isolated in B95a cells, but not in Vero (monkey fibroblast) cells, retain pathogenicity for monkeys^{7,8}. Thus, B95a is now commonly used to isolate MV from clinical specimens. Unlike the Edmonston strain, which is capable of infecting all CD46⁺ primate cell lines, MV strains isolated in B95a or other human B-cell lines infect only some primate B-cell and T-cell lines^{7,9–14}. By using vesicular stomatitis virus (VSV) pseudotypes bearing MV envelope proteins, we showed that virus entry is a principal determinant of cell tropism of MV strains¹⁴. Thus, the receptor molecule that enables B95a-isolated MV strains to enter cells appears to be present on a restricted number of lymphoid cell lines. To identify this receptor, we carried out functional expression cloning using the pseudotype system in which the recombinant VSV containing the green fluorescent

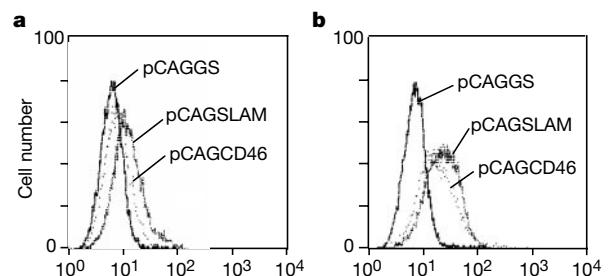


Figure 1 Measles virus binding to CHO cells expressing SLAM or CD46. CHO cells were transiently transfected with pCAGGS, pCAGSLAM or pCAGCD46. The efficiency of transfection was usually more than 50% as determined by flow cytometry analysis. The transfected cells were incubated with the KA (a) or Edmonston (b) strain of MV for 30 min. After washing, the cells were incubated with a monoclonal antibody specific for MV H protein and stained with FITC-labelled anti-mouse IgG.

protein (GFP) in lieu of the VSV G protein (VSVΔG^{*}) was complemented with MV envelope glycoproteins provided in *trans*^{14, 15}.

We assumed that B95a is a good source of messenger RNA encoding a receptor for MV. We generated a cDNA library of B95a cells with the eukaryotic expression vector pCAGGS¹⁶, and divided it into pools each comprising about 450 independent clones. The human kidney cell line 293T, which is susceptible to the Edmonston strain but not to B95a-isolated strains, was transfected with plasmid DNA extracted from individual pools, and infected with VSVΔG^{*} complemented with the haemagglutinin (H) of a B95a-isolated strain KA and the fusion (F) protein of the

Edmonston strain (VSVΔG^{*}-KAHF)¹⁴. The H protein mediates receptor binding, whereas the F protein has membrane-fusion activity². VSVΔG^{*}-KAHF possessed the same tropism as the KA strain¹⁴, and few 293T cells expressed GFP on infection with it. Out of 45 pools of cDNA clones that we tested, a single pool produced significantly more green cells (more than 20 GFP-expressing cells per well by microscope), compared with the other pools (less than 2 GFP-expressing cells per well). This positive pool was further subdivided, and the screening was repeated until a single clone was obtained that could make the transfected 293T cells highly susceptible to VSVΔG^{*}-KAHF (about 10⁴ GFP-expressing cells per well). DNA sequencing revealed that this marmoset cDNA clone

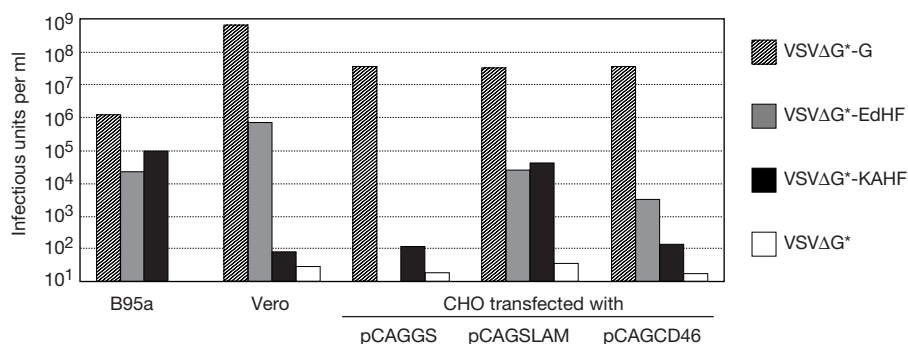


Figure 2 Infectivities of pseudotype viruses on B95a, Vero and transfected CHO cells. CHO cells were used at 40 h after transfection with pCAGGS, pCAGSLAM or pCAGCD46. The cells were infected with VSV pseudotypes (VSVΔG^{*}-G, VSVΔG^{*}-EdHF, VSVΔG^{*}-KAHF

and VSVΔG^{*}), and infectivity titres were determined by counting the number of GFP-expressing cells at 24 h after infection.

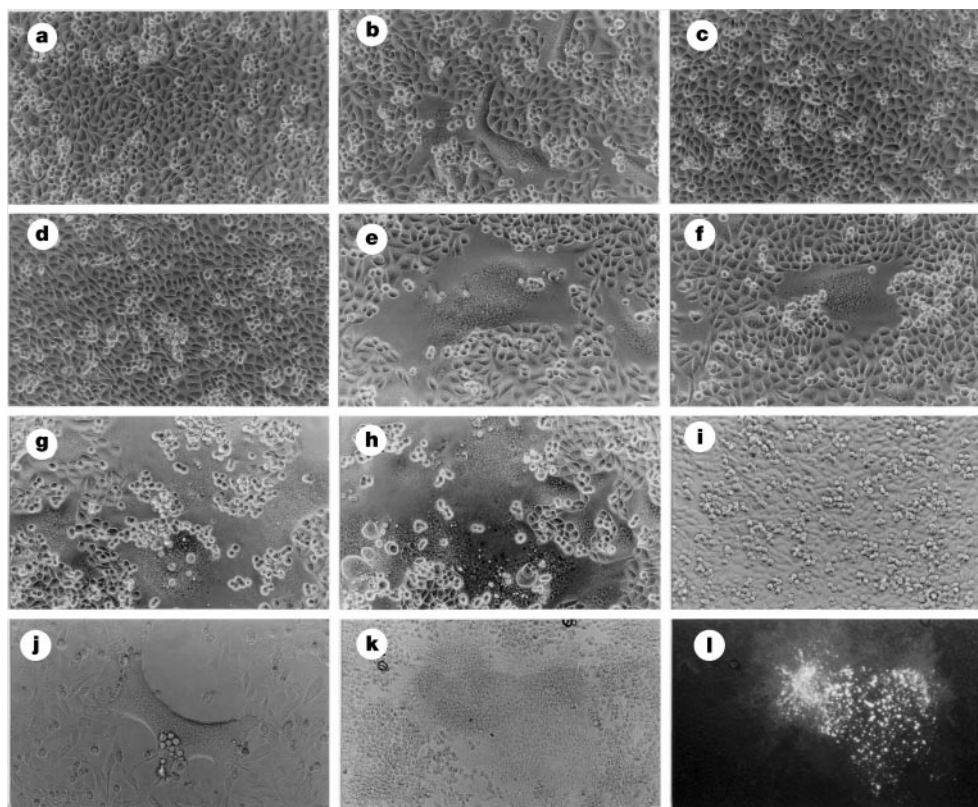


Figure 3 Measles virus infection of CHO cells expressing SLAM or CD46. CHO cells were transiently transfected with pCAGGS (a, d), pCAGSLAM (b, e) or pCAGCD46 (c, f), and then infected with the KA (a–c) or Edmonston (d–f) strain at a multiplicity of infection of 0.5. At 24 h after infection, we observed the cells by microscope. CHO.SLAM cells were either untreated (g), or treated with a mouse IgG1 control antibody (h) or IPO-3 (i), and then infected with the KA strain at a multiplicity of infection of 0.5. The cells were observed

at 24 h after infection. CHO.SLAM cells inoculated with throat swab from a measles patient were observed at 36 h after inoculation (j). In another experiment, CHO.SLAM cells inoculated with throat swab from a measles patient were fixed at 48 h after inoculation, and then stained with anti-MV N protein monoclonal antibody, followed by FITC-labelled anti-mouse IgG. The stained cells were observed under a light microscope (k) or a fluorescence microscope (l).

had a high level of similarity (91% identity at the nucleotide level in the coding region) to the human SLAM gene⁶, suggesting that human SLAM is a cellular receptor for the KA and other MV strains that cannot use CD46 as a receptor. The cDNA clone encoding the membrane-bound form of human SLAM was isolated from peripheral blood mononuclear cells (PBMCs) by polymerase chain reaction with reverse transcription (RT-PCR) based on the published sequence of human SLAM⁶. We confirmed by immunofluorescence staining that this cDNA clone directed the cell-surface expression of SLAM after transfection (data not shown).

We first examined whether MV binds to cells expressing SLAM. Chinese hamster ovary (CHO) cells non-susceptible to MV were transiently transfected with the expression plasmid containing the human SLAM cDNA (pCAGSLAM) or CD46 cDNA (pCAGCD46), and incubated with MV particles. The KA strain exhibited a significant binding to SLAM-expressing CHO cells, but only a low level of binding to CD46-expressing cells, as compared with CHO cells transfected with pCAGGS alone, whereas the Edmonston strain bound well to both SLAM-expressing and CD46-expressing CHO cells (Fig. 1). We then studied the ability of SLAM to allow virus entry into cells by infecting transfected CHO cells with VSV pseudotypes¹⁴: VSVΔG*-G bearing the VSV G protein; VSVΔG*-EdHF bearing the Edmonston H and F proteins; VSVΔG*-KAHF; and VSVΔG* bearing no envelope protein. VSVΔG*-G efficiently infected all of the cells tested, and VSVΔG*-EdHF and VSVΔG*-KAHF exhibited expected infectivities¹⁴ on B95a and Vero cells (Fig. 2). VSVΔG*-KAHF showed more than 100-fold higher titres on SLAM-expressing CHO cells than it did on CHO cells transfected with pCAGGS or pCAGCD46, whereas VSVΔG*-EdHF exhibited more than 100-fold higher titres on SLAM-expressing and CD46-expressing CHO cells than it did on CHO cells transfected with pCAGGS.

CHO cells transfected with pCAGSLAM or pCAGCD46 were also infected with MV at a multiplicity of infection of 1, and virus production was determined at 36 h after infection. On infection with the KA strain, SLAM-expressing CHO cells produced $10^{2.3}$ 50% tissue-culture infectious doses (TCID₅₀) per millilitre of MV, whereas CHO cells transfected with pCAGCD46 or pCAGGS did not produce a detectable level of virus (less than $10^{0.8}$ TCID₅₀ per ml). When infected with the Edmonston strain, virus titre in SLAM-expressing CHO cells ($10^{4.3}$ TCID₅₀ per ml) was 100-fold higher than that in CHO cells transfected with pCAGGS ($10^{2.3}$ TCID₅₀ per ml), and comparable to that in CD46-expressing CHO cells ($10^{4.6}$ TCID₅₀ per ml). In agreement with the results of virus binding, entry and replication, the KA strain caused cytopathic effects (CPEs) in SLAM-expressing CHO cells, but not in CD46-expressing CHO cells (Fig. 3a–c), whereas the Edmonston strain caused CPEs both in SLAM- and in CD46-expressing CHO cells (Fig. 3d–f). Besides CHO cells, murine L cells transfected with pCAGSLAM

produced CPEs on infection with both the KA strain and the Edmonston strain (data not shown). After transfection with pCAGSLAM, primate non-lymphoid HeLa and Vero cells also developed CPEs on infection with the KA strain (data not shown).

We transfected CHO cells with pCAGSLAM plus pCXN2 (ref. 16) containing the *neo* gene, and selected stable clones in the presence of G418. We used the clone expressing the highest level of SLAM (CHO.SLAM) in the following experiments. CHO.SLAM cells developed much stronger CPEs on infection with the KA strain (Fig. 3g) and with the Edmonston strain, as compared with CHO cells transiently expressing SLAM. Using CHO.SLAM cells, we examined whether a monoclonal antibody against human SLAM inhibits MV infection. Cells were pre-incubated with a mouse monoclonal antibody, IPO-3, specific for human SLAM¹⁷ ($10 \mu\text{g ml}^{-1}$) for 1 h, and then infected with MV. (The antibody was present in the medium throughout the experiment.) CHO.SLAM cells untreated or treated with a mouse immunoglobulin (Ig)-G1 control antibody (Fig. 3h) developed CPEs on infection with the KA strain, whereas those treated with IPO-3 did not (Fig. 3i). IPO-3 also inhibited developments of CPEs in CHO.SLAM cells infected with the Edmonston strain and in B95a cells infected with either strain, but did not affect infection of Vero cells with the Edmonston strain (data not shown). We then examined whether six other clinical MV isolates were also able to infect CHO.SLAM cells: four MV strains isolated and propagated in B95a cells that have the same tropism as the KA strain¹³; and two MV strains isolated and propagated in PBMCs that were reported to use CD46 as a receptor¹⁸. Like the KA and Edmonston strains, in CHO.SLAM cells at 24 h after infection, all these MV strains caused CPEs, which were completely blocked by the pre-treatment of the cells with IPO-3 (data not shown).

These results showed that the presence of human SLAM on the cell surface allowed non-susceptible cell lines to bind MV, support MV entry and replication, and develop CPEs, whether the virus was the Edmonston strain, a B95a-isolated strain or a PBMC-isolated strain. In addition, MV infection was inhibited by an antibody against SLAM. We therefore concluded that SLAM is a cellular receptor for MV; the Edmonston strain was shown to use either SLAM or CD46 as a receptor. As the Edmonston strain has been grown in SLAM-negative cells² (and thus should not have adapted to SLAM *in vitro*), the precursor of this strain must have been using SLAM as a receptor in the original patient from which the strain was isolated in 1950s. To confirm the *in vivo* relevance of SLAM, we inoculated CHO.SLAM cells with throat swabs from two measles patients. They developed CPEs at 36 h after inoculation (Fig. 3j). Syncytia were strongly stained with a monoclonal antibody specific for MV nucleoprotein (N)¹⁹ (Fig. 3k, l). Control CHO cells did not develop CPEs. As MV was readily isolated in CHO.SLAM cells without passage, SLAM must be a receptor for MV present in the bodies of these patients.

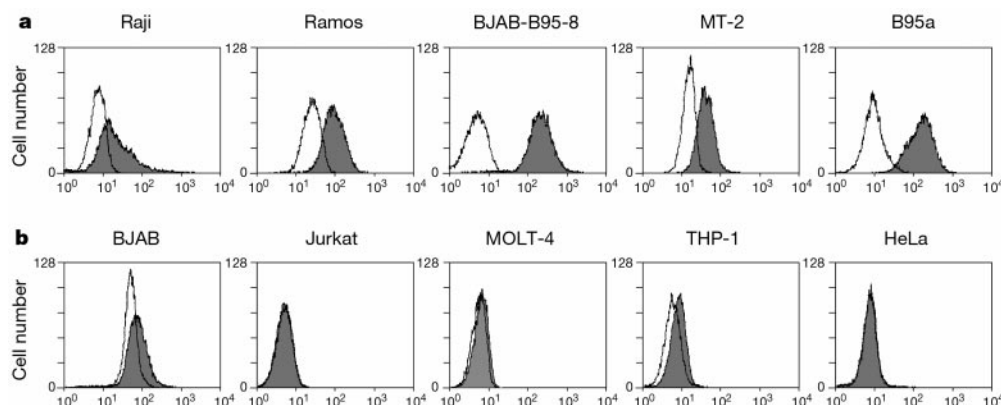


Figure 4 Flow cytometry analysis of the cell-surface expression of SLAM. Cell lines susceptible to B95a-isolated MV strains (**a**) and those non-susceptible (**b**) were incubated

with IPO-3 (solid profiles) or a mouse IgG1 control antibody (empty profiles), followed by staining with FITC-labelled anti-mouse IgG.

We previously showed that human lymphoid cell lines Raji, Ramos, BJAB-B95-8, MT-2, and C91/PL are susceptible to B95a-isolated MV strains, whereas lymphoid cell lines BJAB, Daudi, Jurkat and MOLT-4, myelomonocytic cell lines THP-1, U-937 and HL-60, and adherent cell lines HeLa and Vero are not¹⁴. We carried out immunofluorescence staining to determine whether the expression of SLAM on these cell lines correlates with their susceptibility to B95a-isolated MV strains. All cell lines susceptible to these MV strains expressed SLAM, whereas non-susceptible cell lines had little or no expression of SLAM. Results of representative cell lines were shown in Fig. 4. A high level of SLAM expression on B95a cells would explain why this cell line has been useful in isolating MV from clinical specimens.

SLAM is a glycosylated transmembrane protein that is constitutively expressed on immature thymocytes, CD45RO^{high} memory T cells and a proportion of B cells, and is rapidly induced on T and B cells after activation^{6,17,20,21}. Lymphoid organs (thymus, spleen, lymph nodes) throughout the body are chief sites of MV replication in human patients and experimentally infected monkeys^{2,8,22}. Thus, unlike CD46 which is ubiquitously expressed²³, the tissue distribution of SLAM is more consistent with pathology of MV infection in humans and monkeys. Immunosuppression and lymphopenia characteristic of measles² may be explained by destruction of infected SLAM⁺ lymphocytes. On the other hand, some studies have suggested that MV may inhibit lymphocyte functions without infecting the cell^{24,25}. SLAM is an important costimulatory molecule in T- and B-cell activation^{6,20,21,26–28}. Thus, mere binding of MV particles or envelope proteins to SLAM on the cell surface might affect the signals induced through SLAM, thereby impairing lymphocyte activation. It is noteworthy that the engagement of SLAM by a monoclonal antibody induces Th0/Th1 cytokine production profile²¹, and there is both *in vivo* and *in vitro* evidence of Th2 polarization in cytokine responses during and after measles²⁹. Measles virus is a member of the *Morbillivirus* genus that also includes canine distemper virus and rinderpest (cattle plague) virus. These viruses cause devastating animal diseases that have many similarities to measles, including immunosuppression². Furthermore, they replicate well in SLAM⁺ B95a cells. From these observations, we propose that destruction of SLAM⁺ cells and/or virus–SLAM interaction may be a principal mechanism of the immunosuppressive nature of morbilliviruses. To test this idea, we are now determining whether SLAM also acts as a receptor for other members of morbilliviruses. □

Methods

Viruses and cells

We grew and titrated the Edmonston strain of MV (American Type Culture Collection) on Vero cells. The MV strains isolated from patients with measles by using B95a cells between 1984 and 1996 were provided by F. Kobune^{7,13}. They were grown and titrated on B95a cells. The MV strains isolated in PBMCs (JW and IV) were provided by M. Manchester¹⁸. Viruses used for binding analysis were purified by sucrose gradient centrifugation. The preparation of the pseudotype viruses (VSVΔG*-KAHF, VSVΔG*-EdHE, VSVΔG*-G and VSVΔG*) and derivations and culture conditions of cell lines used have been described¹⁴.

Expression cloning

Poly(A)⁺ RNA from B95a cells was used to synthesize cDNA with the Superscript plasmid system (Life Technologies). cDNA molecules larger than 500 base-pairs long were recovered from agarose gel and ligated unidirectionally to pCAGGS digested with *Eco*RI and *Not*I (pCAGGS has been modified to contain the multicloning site). The transformed cells were divided into 210 pools, and individual pools were cultured at 37°C overnight. The total number of independent clones (9.4×10^4) was determined by plating a portion of the transformed cells. Thus, each pool contained about 450 independent clones. Plasmid DNA was extracted from each pool. 293T cells (1×10^5 cells per well) were plated in 24-well plates, and transfected with 0.25 µg of plasmid DNA from each pool by using Lipofectamine (Life Technologies). At 24–48 h after transfection, we infected 293T cells with 1×10^5 infectious units of VSVΔG*-KAHF (titrated on B95a cells), and incubated them at 37°C in a CO₂ incubator. At 14–24 h after infection, we determined susceptibility to the pseudotype virus by counting the number of GFP-expressing cells under a fluorescence microscope.

Isolation of the human SLAM cDNA

The human SLAM cDNA was obtained by reverse transcription of total RNA from phytohaemagglutinin-stimulated human PBMCs followed by polymerase chain reaction (PCR). Primers used for the PCR were 5'-CCCCAATTC AGACAGCTCTGCTG CATGAC-3' and 5'-AAAGCGGCGCCCTTCAGAAAGTCCC TTTGTTGG-3'. Sequences underlined are sites for restriction enzymes. The PCR product was subcloned into pCAGGS and sequenced. The cDNA clone encoding the membrane-bound form of SLAM with the complete cytoplasmic tail⁶ was selected for the experiments.

Binding analysis

CHO cells (5×10^5 cells) were transfected with 1 µg of pCAGGS containing human SLAM cDNA (pCAGSLAM) or human CD46 cDNA (C2 isoform; pCAGCD46) or pCAGGS alone by using Lipofectamine. At 24 h after transfection, we collected the CHO cells by treating them with PBS containing 1 mM EDTA. The cells were mixed with purified viruses (1×10^5 TCID₅₀), and incubated in 100 µl of PBS containing 3% fetal bovine serum at 4°C for 30 min. After being washed three times with PBS, the cells were incubated with a monoclonal antibody against MV H protein³⁰, followed by staining with fluorescein isothiocyanate (FITC)-labelled goat anti-mouse IgG. The stained cells were analysed on a FACScan machine (Becton Dickinson). Dead cells that positively stained with propidium iodide were excluded from the analysis.

Immunofluorescence staining

CHO.SLAM cells inoculated with throat swab were fixed with acetone:methanol (1:1), and stained with anti-MV N protein monoclonal antibody essentially as described¹⁹. Cell lines were stained with IPO-3 (Kamiya Biomedical) or a mouse IgG1 control antibody and then with FITC-labelled goat anti-mouse IgG. Flow cytometry analysis was carried out as described above.

Received 12 May; accepted 10 June 2000.

- Oldstone, M. B. A. *Viruses, Plagues, & History* (Oxford Univ. Press, New York, 1998).
- Griffin, D. E. & Bellini, W. J. in *Fields Virology* (eds. Fields, B. N. et al.) 1267–1312 (Lippincott-Raven, Philadelphia, 1996).
- Naniche, D. et al. Human membrane cofactor protein (CD46) acts as a cellular receptor for measles virus. *J. Virol.* **67**, 6025–6032 (1993).
- Dorig, R. E., Marciel, A., Chopra, A. & Richardson, C. D. The human CD46 molecule is a receptor for measles virus (Edmonston strain). *Cell* **75**, 295–305 (1993).
- Buckland, R. & Wild, T. F. Is CD46 the receptor for measles virus? *Virus Res.* **48**, 1–9 (1997).
- Cocks, B. G. et al. A novel receptor involved in T-cell activation. *Nature* **376**, 260–263 (1995).
- Kobune, F., Sakata, H. & Sugiyama, A. Marmoset lymphoblastoid cells as a sensitive host for isolation of measles virus. *J. Virol.* **64**, 700–705 (1990).
- Kobune, F. et al. Nonhuman primate models of measles. *Lab. Anim. Sci.* **46**, 315–320 (1996).
- Schneider-Schaulies, J. et al. Receptor usage and differential downregulation of CD46 by measles virus wild-type and vaccine strains. *Proc. Natl Acad. Sci. USA* **92**, 3943–3947 (1995).
- Schneider-Schaulies, J., Dunster, L. M., Kobune, F., Rima, B. & Ter Meulen, V. Differential downregulation of CD46 by measles virus strains. *J. Virol.* **69**, 7257–7259 (1995).
- Lecouturier, V. et al. Identification of two amino acids in the hemagglutinin glycoprotein of measles virus (MV) that govern hemadsorption, HeLa cell fusion, and CD46 downregulation: phenotypic markers that differentiate vaccine and wild-type MV strains. *J. Virol.* **70**, 4200–4204 (1996).
- Hsu, E. C. et al. A single amino acid change in the hemagglutinin protein of measles virus determines its ability to bind CD46 and reveals another receptor on marmoset B cells. *J. Virol.* **72**, 2905–2916 (1998).
- Tanaka, K., Xie, M. & Yanagi, Y. The hemagglutinin of recent measles virus isolates induces cell fusion in a marmoset cell line, but not in other CD46-positive human and monkey cell lines, when expressed together with the F protein. *Arch. Virol.* **143**, 213–225 (1998).
- Tatsuo, H. et al. Virus entry is a major determinant of cell tropism of Edmonston and wild-type strains of measles virus as revealed by vesicular stomatitis virus pseudotypes bearing their envelope proteins. *J. Virol.* **74**, 4139–4145 (2000).
- Takada, A. et al. A system for functional analysis of Ebola virus glycoprotein. *Proc. Natl Acad. Sci. USA* **94**, 14764–14769 (1997).
- Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants by a novel eukaryotic vector. *Gene* **108**, 193–200 (1991).
- Sidorenko, S. P. & Clark, E. A. Characterization of a cell surface glycoprotein IPO-3, expressed on activated human B and T lymphocytes. *J. Immunol.* **151**, 4614–4624 (1993).
- Manchester, M. et al. Clinical isolates of measles virus use CD46 as a cellular receptor. *J. Virol.* **74**, 3967–3974 (2000).
- Huber, M. et al. Measles virus phosphoprotein retains the nucleocapsid protein in the cytoplasm. *Virology* **185**, 299–308 (1991).
- Aversa, G. et al. SLAM and its role in T cell activation and Th cell responses. *Immunol. Cell Biol.* **75**, 202–205 (1997).
- Aversa, G., Chang, C.-C., Carballido, J. M., Cocks, B. G. & de Vries, J. E. Engagement of the signaling lymphocytic activation molecule (SLAM) on activated T cells results in IL-2-independent, cyclosporin A-sensitive T cell proliferation and IFN-γ production. *J. Immunol.* **158**, 4036–4044 (1997).
- McChesney, M. B. et al. Experimental measles. I. Pathogenesis in the normal and the immunized host. *Virology* **233**, 74–84 (1997).
- Lisowski, M. K., Post, T. W. & Atkinson, J. P. Membrane cofactor protein (MCP or CD46): newest member of the regulators of complement activation gene cluster. *Annu. Rev. Immunol.* **9**, 431–455 (1991).
- Yanagi, Y., Cubitt, B. A. & Oldstone, M. B. A. Measles virus inhibits mitogen-induced T cell proliferation but does not directly perturb the T cell activation process inside the cell. *Virology* **187**, 280–289 (1992).

25. Schlender, J. *et al.* Interaction of measles virus glycoproteins with the surface of uninfected peripheral blood lymphocytes induces immunosuppression in vitro. *Proc. Natl Acad. Sci. USA* **93**, 13194–13199 (1996).
26. Punnonen, J. *et al.* Soluble and membrane-bound forms of signaling lymphocytic activation molecule (SLAM) induce proliferation and Ig synthesis by activated human B lymphocytes. *J. Exp. Med.* **185**, 993–1004 (1997).
27. Sayos, J. *et al.* The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* **395**, 462–469 (1998).
28. Mihalap, S. V. *et al.* CDw150 associates with Src-homology 2-containing inositol phosphatase and modulates CD95-mediated apoptosis. *J. Immunol.* **162**, 5719–5727 (1999).
29. Griffin, D. E. & Ward, B. J. Differential CD4 T cell activation in measles. *J. Infect. Dis.* **168**, 275–281 (1993).
30. McFarlin, D. E., Bellini, W. J., Mingioli, E. S., Behar, T. N. & Trudgett, A. Monospecific antibody to the haemagglutinin of measles virus. *J. Gen. Virol.* **48**, 425–429 (1980).

Acknowledgements

We thank M. A. Whitt for allowing us to use the VSVΔG⁺-GFP system; F. Kobune and M. Manchester for providing MV strains; T. Aoki for providing throat swabs from measles patients; and Y. Matsuura, K. Ishihara and members of our laboratory for helpful discussions. This work was supported by grants from the Ministry of Education, Science and Culture of Japan and from the Program for Promotion of Fundamental Studies in Health Sciences of the Organization for Drug ADR Relief, R&D Promotion and Product Review of Japan.

Correspondence and requests for materials should be addressed to Y.Y. (e-mail: yyanagi@virology.med.kyushu-u.ac.jp). The GenBank accession number for B95a SLAM cDNA sequence is AF257239.

Colorectal carcinomas in mice lacking the catalytic subunit of PI(3)K γ

Takehiko Sasaki*, Junko Irie-Sasaki*, Yasuo Horie, Kurt Bachmaier*, Jimmie E. Fata†, Martin Li‡, Akira Suzuki*, Dennis Bouchard*, Alexandra Ho*, Mark Redston§, Steven Gallinger§, Rama Khokha†, Tak W. Mak*, Phillip T. Hawkins||, Len Stephens||, Stephen W. Scherer‡, Ming Tsao¶ & Josef M. Penninger*

*Amgen Institute, Ontario Cancer Institute, and the Departments of Medical Biophysics and Immunology, University of Toronto, 620 University Avenue, Toronto, Ontario M5G 2C1, Canada

†Department of Medical Biophysics and Department of Laboratory Medicine and Pathobiology, Ontario Cancer Institute, University of Toronto, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada

‡Department of Genetics and Genomic Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

§Centre for Cancer Genetics, Samuel Lunenfeld Research Institute, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

||The Babraham Institute, Babraham, Cambridge CB2 4AT, UK

¶Departments of Laboratory Medicine and Pathology, University of Toronto, Toronto, 600 University Avenue, Ontario M5G 1X5, Canada

Phosphoinositide-3-OH kinases (PI(3)Ks) constitute a family of evolutionarily conserved lipid kinases that regulate a vast array of fundamental cellular responses, including proliferation, transformation, differentiation and protection from apoptosis^{1,2}. PI(3)K-mediated activation of the cell survival kinase PKB/Akt, and negative regulation of PI(3)K signalling by the tumour suppressor PTEN (refs 3, 4) are key regulatory events in tumorigenesis^{5–7}. Thus, a model has arisen that PI(3)Ks promote development of cancers. Here we report that genetic inactivation of the p110 γ catalytic subunit of PI(3)K γ (ref. 8) leads to development of invasive colorectal adenocarcinomas in mice. In humans, p110 γ protein expression is lost in primary colorectal adenocarcinomas from patients and in colon cancer cell lines. Overexpression of wild-type or kinase-dead p110 γ in human colon cancer cells with mutations of the tumour suppressors APC and p53, or the oncogenes β -catenin and Ki-ras, suppressed tumorigenesis.

Thus, loss of p110 γ in mice leads to spontaneous, malignant epithelial tumours in the colorectum and p110 γ can block the growth of human colon cancer cells.

Four mammalian type 1 PI(3)K catalytic isoforms (p110 α , β , γ , δ) have been identified¹. Among them, p110 γ is a unique member of the enzymes. The type 1A PI(3)Ks, p110 α , p110 β and p110 δ , associate with the p85 family of regulatory subunits, but type 1B p110 γ binds to a p101 adaptor molecule^{9,10}. Whereas type 1A PI(3)Ks are activated by interaction with tyrosine-phosphorylated molecules, p110 γ is activated by engagement of G-protein-coupled receptors (GPCRs). Recently, we⁸ and others^{11,12} showed that mice homozygous for a gene-targeted deletion of the p110 γ gene are born at the expected mendelian ratio and appear healthy up to two months old, indicating that p110 γ is not required for embryogenesis or organogenesis. In those studies, we showed that p110 γ is an important regulator of neutrophil migration and T-cell functions.

We now report that p110 γ ^{−/−} mice have shorter lifespans than their heterozygous littermates (Fig. 1a). While there are no abnormalities in appearance, size, body weight or health of young p110 γ ^{−/−} mice (< 8 weeks), older p110 γ ^{−/−} mice have an increased incidence of premature death associated with progressive weight loss (Fig. 1b), development of macroscopically visible tumours associated with swollen abdomens, dilated large bowels suggestive of colonic obstructions and rectal prolapses (Fig. 1c). This phenotype was observed in each of two independent lines of p110 γ ^{−/−} mice derived from germline transmission of two distinct embryonic stem cell clones carrying the mutated allele. Similarly to human colorectal cancer¹³, most tumours in p110 γ ^{−/−} mice occurred at the proximal and distal parts of the large intestine. Grossly evident colorectal tumours appeared first in 12-week-old animals but subsequently increased in frequency so that about 40% of p110 γ ^{−/−} mice had at least one visible tumour (> 5 mm) by 6 months of age (see Table 1 in Supplementary Information). The progressive weight loss is probably due to malnutrition secondary to the colonic tumours.

Histological analysis revealed that the colonic tumours in p110 γ ^{−/−} mice were epithelial in origin and represented all stages of transformation. We observed hyperplastic and adenomatous proliferation, including tubular and villous adenomas (Fig. 1e, f), as well as invasive adenocarcinomas that arose from these adenomas (Fig. 1g–i). Larger carcinomas demonstrated transmural, local peritoneal invasion, and metastasis into the peritoneal cavity. We also noted adenomatous polyps with varying degrees of dysplasia. In some mutant mice, the colonic adenomas and carcinomas were localized to defined areas, whereas other mice showed multiple adenomas throughout the length of the large bowel. We found no tumours in the small intestine, stomach, or any other epithelial tissues. p110 γ ^{−/−}rag1^{−/−} mice, which have no T or B lymphocytes, also had spontaneous colorectal carcinomas. It remains to be determined whether all p110 γ ^{−/−} mice will eventually develop colon tumours, or whether some will be spared, owing to the presence of modifying genes and/or environmental influences¹⁴. These results show that loss of p110 γ expression in mice results in the development of spontaneous, malignant epithelial tumours in the colorectum.

The p110 γ gene was originally cloned from haematopoietic cells^{9,10}. However, p110 γ messenger RNA is also expressed in heart, skeletal and smooth muscle, lung, liver, kidney, brain, prostate, exocrine pancreas, salivary glands, parotid gland, testis, uterus and intestine in adult mice (ref. 15 and data not shown). p110 γ protein and mRNA are expressed in the colon as determined by western blotting and *in situ* RNA hybridization (see Fig. 1 in Supplementary Information). In p110 γ ^{−/−} mice, we observed no increase in expression of the type 1A PI(3)Ks, p110 α and p110 β , and no changes in the expression levels of the regulatory p85 subunit of type 1A PI(3)Ks or PKB. To analyse further changes in type 1A PI(3)K activity in the colon, we determined PI(3,4,5)-trisphosphate production in p85 immunoprecipitates. p85-asso-

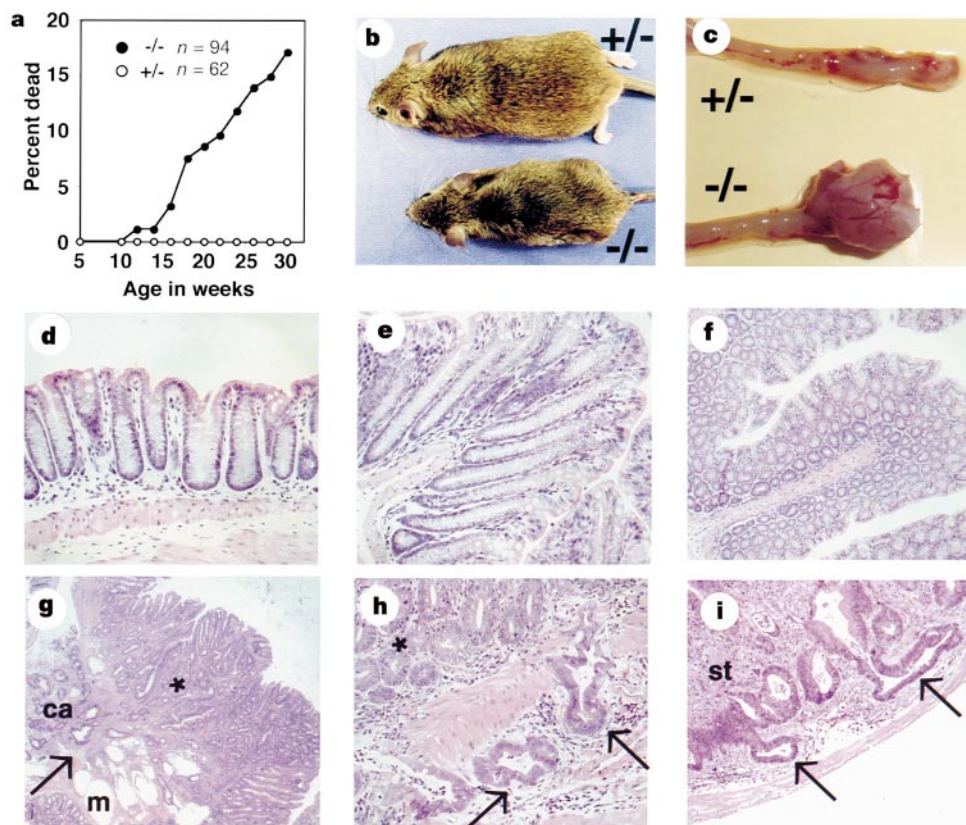


Figure 1 Formation of colorectal adenocarcinomas in $p110\gamma^{-/-}$ mice. **a**, Mortality of $p110\gamma^{+/+}$ and $p110\gamma^{-/-}$ mice. **b**, **c**, Gross appearance (**b**) and large tumour (**c**) in colorectum of a $p110\gamma^{-/-}$ mouse. **d**, Normal colonic mucosa. **e**, Hyperplastic lesion showing increased crypt thickness. **f**, Tubular adenoma in ascending colon of 14-week-old $p110\gamma^{-/-}$ mouse. **g**, Adenocarcinoma (ca) that has transmurally invaded into the submucosa and muscularis propria (arrow) arising in a tubulovillous adenoma (asterisk).

Extracellular mucin pools (m) are evident. **h**, Adenocarcinoma arising in a rectal tubular adenoma (asterisk). Neoplastic glands (arrows) have infiltrated into and through the muscularis mucosa. **i**, Adenocarcinoma that has invaded into the submucosa and muscularis propria in the rectum of a 20-week-old $p110\gamma^{-/-}$ mouse. Note irregular shapes of malignant glands (arrows) and desmoplastic stroma (st) associated with invasive tumour. Magnifications: **d**, **e**, $\times 50$; **f**, **g**, $\times 20$; **h**, **i**, $\times 128$.

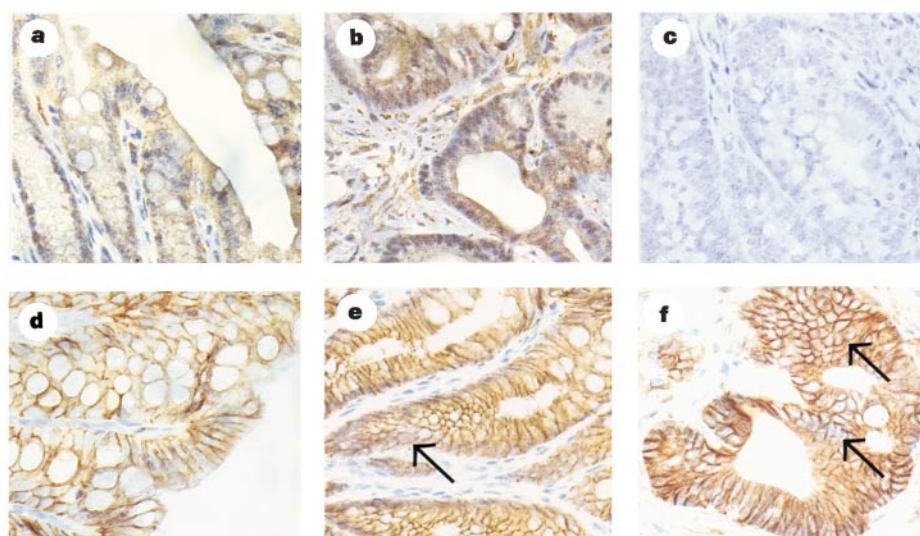


Figure 2 APC and β -catenin expression in neoplastic lesions. **a–c**, Immunohistochemical determination of APC expression using an anti-APC antibody reactive to 20 amino acids at the C terminus. **a**, Normal $p110\gamma^{+/+}$ epithelial cells. **b**, Transformed $p110\gamma^{-/-}$ cells. **c**, Negative control using an isotype matched antibody. **d–f**, Immunohistochemical determination of β -catenin. Staining of colonic epithelial cells

in normal $p110\gamma^{+/+}$ colon (**d**), $p110\gamma^{-/-}$ villous adenoma (**e**) and $p110\gamma^{-/-}$ invasive adenocarcinoma (**f**). No differences in expression levels or cellular localization of APC or β -catenin were observed in colons from $p110\gamma^{-/-}$ mice. β -catenin was expressed in the cytoplasm of tumour cells but did not accumulate in the nucleus (arrows). Magnifications: **a**, **d**, **f**, $\times 256$; **b**, **c**, **e**, $\times 200$.

ciated type 1A PI(3)K activities are comparable among $p110\gamma^{+/-}$ and two different $p110\gamma^{-/-}$ mice (see Fig. 1 in Supplementary Information). Thus, loss of $p110\gamma$ does not apparently lead to upregulation of type 1A PI(3)K protein or altered type 1A PI(3)K activity.

Tumour progression from normal colonic epithelium to early-

and late-stage adenomas and carcinomas occurs in defined stages. These stages are controlled by the inactivation of tumour suppressor genes such as Adenomatous Polyposis Coli (APC), p53, SMADs or the DNA repair genes MSH2 and MLH1, in coordination with the activation of oncogenes, including β -catenin, CyclinD1, Ras, Bcl-2,

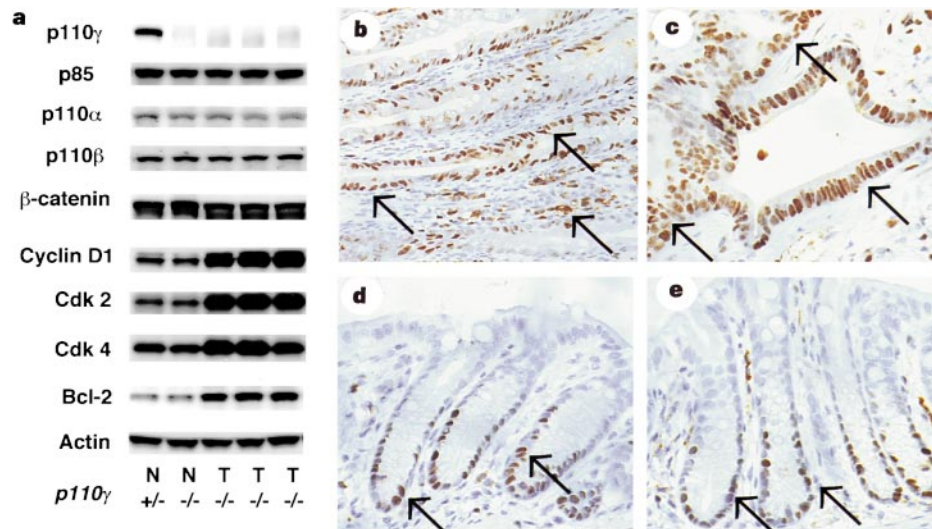


Figure 3 Upregulation of cell-cycle molecules and increased proliferation in $p110\gamma^{-/-}$ tumours. **a**, Expression levels of proteins in the normal epithelium of one $p110\gamma^{+/-}$ mouse (N, +/–), normal epithelium of one $p110\gamma^{-/-}$ mouse (N, –/–) and adenocarcinomas from three different $p110\gamma^{-/-}$ mice (T, –/–). Similar results were obtained using normal epithelium of additional $p110\gamma^{+/-}$ mice ($n = 5$), normal epithelium of $p110\gamma^{-/-}$ mice ($n = 5$) and adenocarcinomas from $p110\gamma^{-/-}$ mice ($n = 7$). **b–e**, Proliferation was

assessed by *in situ* staining for expression of the proliferating-cell nuclear antigen (PCNA). All tumour cells in villous adenomas (**b**) and adenocarcinomas (**c**) in $p110\gamma^{-/-}$ mice have entered the cell cycle and express PCNA. In normal tissue of $p110\gamma^{-/-}$ (**d**) and $p110\gamma^{+/-}$ (**e**) mice, only the epithelial cells in crypts show positive PCNA staining. Magnifications: **b, d, e**, $\times 200$; **c**, $\times 100$.

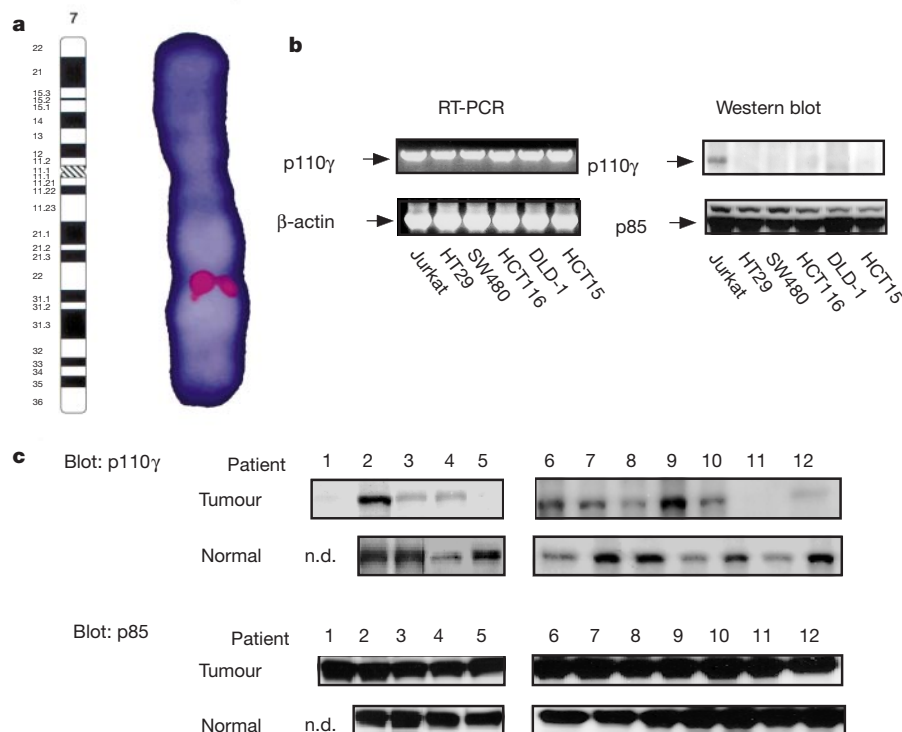


Figure 4 Decreased expression of $p110\gamma$ protein in human colon cancer cell lines and primary colon cancers from patients. **a**, FISH mapping of the human $p110\gamma$ gene to chromosome 7q22. The $p110\gamma$ gene is shown in red and the chromosome was counterstained with DAPI (blue). **b**, RT–PCR analysis for mRNA expression and western blotting for $p110\gamma$ protein expression in human colon cancer cell lines. Antibodies were reactive to amino-acid residues 2–17 of human $p110\gamma$. Absence of $p110\gamma$ protein was

confirmed with two other antibodies reactive to residues 742–756 and 1,032–1,050. Human Jurkat cells are shown as control. Absence of $p110\gamma$ protein was also observed in the human colon cancer cell lines LIM1215, Caco2 and LIM2099. **c**, Western blotting for $p110\gamma$ and p85 expression in normal colon epithelium (normal) and adenocarcinomas (tumour) isolated from 12 different patients. n.d., not determined.

Myc and Cdk2 (reviewed in refs 13, 16, 17). To determine the potential cooperation between p110 γ and known oncogenes and tumour suppressors, we first examined the expression of APC and β -catenin in colonic tissues by immunohistochemistry. Functional APC targets β -catenin for degradation, suppressing β -catenin-mediated gene expression and tumorigenesis^{18–20}. Normal levels of APC expression were detected with antibodies reactive to carboxy- (Fig. 2a, b) and amino-terminal (not shown) regions of APC in both adenocarcinomas and normal epithelium of p110 γ ^{−/−} mice. Furthermore, there were no significant differences in the expression or subcellular localization of β -catenin in p110 γ ^{−/−} normal epithelium, adenomas and carcinomas (Fig. 2d–f). It therefore appears that neither the loss (or truncation) of APC, nor accumulation of β -catenin, is a primary cause for the development of colorectal cancers in mice lacking p110 γ .

We next prepared protein extracts from normal colon of p110 γ ^{+/−} mice, normal non-transformed colon of p110 γ ^{−/−} mice and colonic

tumours of p110 γ ^{−/−} mice, to examine the expression of proteins implicated in colon cancer. Consistent with the immunohistological analysis, we observed no significant difference in the amount of β -catenin protein among normal and transformed epithelial tissue (Fig. 3a). Also, we detected no apparent differences in the expression of c-Myc, SMAD1, SMAD2, SMAD3, SMAD4, p53, p21, p27, FasL, Bcl-XL or MSH2 (not shown). However, expression of the death suppressing molecule Bcl-2 was upregulated in all p110 γ ^{−/−} tumours analysed (Fig. 3a). In addition, expression of the cell-cycle molecules Cyclin D1, Cdk2 and Cdk4 was substantially higher in p110 γ ^{−/−} adenocarcinomas than in normal colonic epithelial cells (Fig. 3a). Expression of the p85 regulatory subunit of the type 1A PI(3)Ks (Fig. 3a), the type 1A catalytic PI(3)K isoforms p110 α and p110 β , and Cyclin A and Cyclin E (not shown) in the colon was not affected by the loss of p110 γ . Consistent with the observed upregulation of Cyclin D1, Cdk2 and Cdk4, cellular hyperproliferation was present throughout p110 γ ^{−/−} adenomas and adenocarcinomas as assessed by *in situ* immunostaining for proliferating cell nuclear antigen (PCNA; Fig. 3b, c). Thus, p110 γ ^{−/−} colorectal cancers display increased expression of molecules involved in cell-cycle progression and protection from cell death.

Our data show that mice lacking p110 γ develop spontaneous multifocal adenomas and invasive adenocarcinomas in the colon, indicating that p110 γ suppresses colorectal cancer *in vivo*. To extend these findings to human colorectal cancer, we first analysed the chromosomal localization of the p110 γ gene and p110 γ expression in primary adenocarcinomas from patients and human colon cancer cell lines. Using fluorescence *in situ* hybridization (FISH) with sequences spanning the whole human p110 γ gene (PIK3CG), we placed the gene at the position 7cen-EPO-CUTL1-RELN-ORC5L-PIK3CG-PRKAR2B-PDS-DRA within the chromosome 7q22 band (Fig. 4a; see also <http://www.genet.sickkids.on.ca/chromosome7>). In five human colon cancer cell lines, HT29, SW480, HCT116, DLD-1 and HCT-15, there was no loss of heterozygosity (LOH) at the p110 γ locus and these cells still expressed p110 γ mRNA (Fig. 4b). However, this does not exclude the possibility that some primary human colon cancers display LOH at the p110 γ locus and/or contain p110 γ gene mutations. Surprisingly, none of the eight human cancer cell lines analysed expressed any detectable p110 γ protein (Fig. 4b). Thus, p110 γ expression might be regulated by protein degradation. Whereas p110 γ protein can readily be detected in normal colonic mucosal tissue, the expression of p110 γ protein was not detectable in 3 out of 12 primary adenocarcinomas isolated from the colons of human patients (Fig. 4c). The frequency and mechanism(s) of ablated p110 γ protein expression in primary colorectal tumours need to be determined.

To test whether p110 γ can directly influence the growth and tumorigenic ability of epithelial colon cancer cells, we overexpressed wild-type or kinase-dead p110 γ in the human colon cancer cell lines HCT116, DLD-1 and HT29 (ref. 21). HCT116 has activating mutations in β -catenin and Ki-ras, but normal APC and p53. DLD-1 has mutations in the tumour suppressors APC and p53, activated Ki-ras, but normal β -catenin. HT29 has mutations in APC and p53, but normal Ki-ras and β -catenin. Ectopic expression of wild-type p110 γ caused a marked inhibition of *in vitro* colony formation in all three colon cancer cell lines (Fig. 5a). Moreover, *in vivo* tumour formation of HCT116 cells in nude mice was impaired in cells expressing wild-type p110 γ as compared with HCT116 cells transfected with a control vector (Fig. 5b, c). Surprisingly, a kinase-dead mutant of p110 γ also suppressed *in vitro* colony formation of HCT116, DLD-1 and HT29, and *in vivo* tumorigenesis of HCT116 cells to a comparable extent as wild-type p110 γ . Thus, the enzymatic PI(3)K activity is dispensable for the growth-inhibitory and anti-tumorigenic effects of p110 γ on colon cancer cell lines. These results indicate that expression of p110 γ has direct effects on the *in vitro* and *in vivo* growth of human colon cancer cells that carry mutations in the tumour suppressors APC and p53 or activating

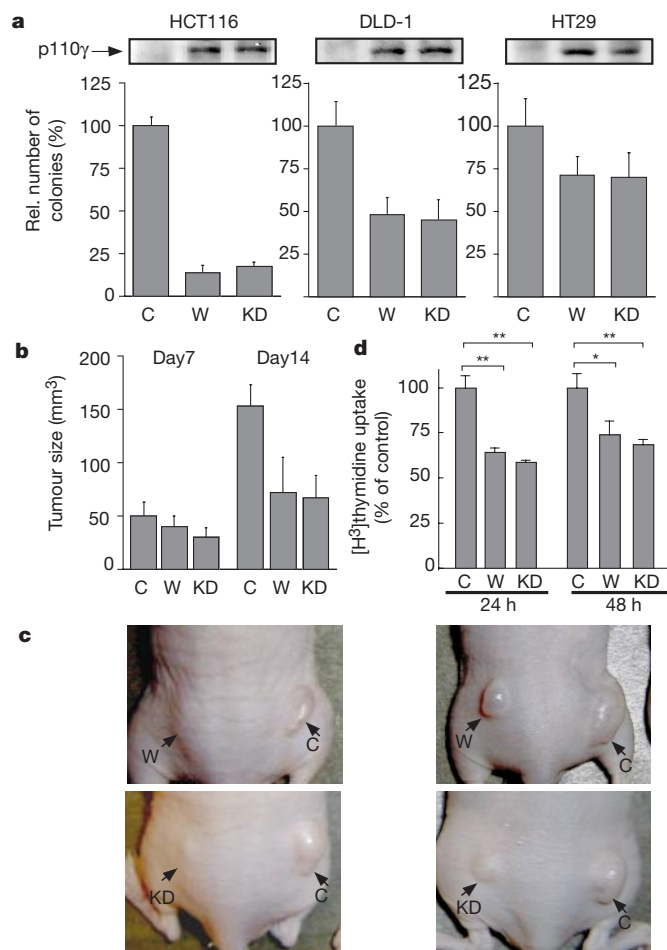


Figure 5 Overexpression of p110 γ suppresses *in vitro* colony formation and *in vivo* tumour growth of human colon cancer cell lines. **a**, The colon cancer cell lines HCT116, DLD-1 and HT29 were transfected with a control vector (C), wild-type (W) or kinase-dead (KD) p110 γ . Numbers of G418-resistant colonies ($\pm$ s.d.) were scored after 2 weeks of selection. Western blots for p110 γ expression are shown for each cell line. **b**, **c**, p110 γ impairs *in vivo* tumorigenesis of HCT116 cells. **b**, Mean ($\pm$ s.e.m.) tumour sizes (mm³) were measured 7 and 14 days after subcutaneous injection of control HCT116 cells, injected into the right flanks, and HCT116 cells overexpressing either wild-type or kinase-dead p110 γ , injected into the left flank of the same athymic nu/nu mice. **c**, Representative photographs of tumour growth in nu/nu mice 18 days after subcutaneous injection of HCT116 transfectants (see **b**). **d**, HCT116 cells transfected with a control vector, wild-type or kinase-dead p110 γ were activated with 5% serum. Mean DNA synthesis ($\pm$ s.e.m.) was measured by [³H]thymidine incorporation. Statistical differences (*t*-test) were *P* < 0.05 (asterisk) or *P* < 0.01 (two asterisks).

mutations of the oncogenes *Ki-ras* and β -catenin.

Colon cancers isolated from *p110 γ ^{-/-}* mice displayed increased expression of the death inhibitor Bcl-2 and the cell-cycle regulatory molecules Cyclin D1, Cdk2 and Cdk4 (Fig. 3). Thus, we analysed whether expression of p110 γ in human colon cancer cells would affect proliferation and/or apoptosis. Expression of either wild-type or kinase-dead p110 γ in HCT116 cells suppressed DNA synthesis (Fig. 5d) and partially blocked G1–S phase cell-cycle progression (see Fig. 2a in Supplementary Information), without any significant changes in cell viability. The extent and kinetics of apoptosis in response to serum withdrawal or sodium butyrate as well as expression levels of Bcl-2 were comparable among HCT116 cells transfected with a control vector and HCT116 overexpressing wild-type or kinase-dead p110 γ (not shown). Expression of wild-type or kinase-dead p110 γ had no apparent effect on the protein amount of β -catenin, p85, p110 α or p110 β . However, although baseline Cyclin D1 expression was not affected, serum-induced Cyclin D1 expression was markedly impaired in HCT116 cells expressing wild-type or kinase-dead p110 γ (Fig. 2b in Supplementary Information). These data show that p110 γ can inhibit cell-cycle progression in the colon cancer cell line HCT116.

Colon cancers are the second leading cause of cancer death and it has been estimated that 50% of humans develop colon tumours by age 70 (ref. 13). The incidence of colon cancer correlates with environmental factors such as diet²². It has previously been shown that mice deficient in the α_{12} gene develop adenocarcinomas of the colon in the absence of overt tumours in any other organs²³. Taken in the context of our findings with *p110 γ ^{-/-}* mice, we propose that a α_{12} -coupled receptor(s) is expressed on colonic epithelial cells, and that p110 γ interacts with this receptor to negatively regulate multiple steps of tumour progression. Identification of such a GPCR and downstream signalling pathway that control differentiation and proliferation of intestinal epithelial cells may be helpful in the design of novel strategies for colon cancer treatment. Our data lead to the conclusion that the PI(3)K isoform, p110 γ , can inhibit the growth of colon cancer cells and suppresses the development of colorectal cancers *in vivo*. □

Methods

Mice

p110 γ mutant mice have been described⁸. Athymic *nu/nu* mice and *rag1* knockout mice were purchased from the Jackson Laboratory. Genotypes were determined by Southern blotting using genomic DNA. Two different embryonic stem cell lines were used to generate a large cohort of the mutant mice. The mice belonged to generations F₂–F₄ backcrossed to C57BL/6J. All mice were maintained at the animal facilities of the Ontario Cancer Institute according to institutional guidelines.

Colon cancer cell lines, human colorectal tumour samples and genomic FISH

The following colon cancer cell lines were used: HT29, SW480, HCT116, DLD-1, HCT15, LIM1215, CaCo2, LIM2099. All cells were cultured in RPMI medium containing 10% fetal calf serum (FCS). Human colorectal cancer tissues were collected from surgically resected colorectal specimens of patients following approval by the institutional ethic committee. Normal non-transformed mucosal tissue was collected from during colectomy after stripping the mucosal tissue from the muscle layer of the bowel wall. Genomic FISH analysis was performed using cosmid probe cos248c8 that contains exons 1–8 of the human *p110 γ* gene. The human *p110 γ* gene contains 11 exons and encompasses 42 kilobases of genomic DNA contained in a bacterial artificial chromosome clone GS223D04 (GenBank accession number AC005018). Polymerase chain reaction followed by reverse transcription (RT–PCR) was performed, amplifying a 449-base-pair fragment of *p110 γ* complementary DNA.

Histology, immunohistochemistry and *in situ* hybridization

Intestines were isolated and opened longitudinally to inspect for mucosal tumours. Tissue samples were fixed in 4% buffered formalin and embedded in paraffin. 5- μ m sections were cut and stained with haematoxylin and eosin. For immunohistochemistry, paraffin-embedded sections were dehydrated and incubated for 10 min at 95 °C in 10 mM citrate buffer (pH 6.0) to expose antigenic epitopes. Sections were then incubated with the following antibodies: anti-APC (C terminus; C-20; Santa Cruz Biotech; 1:200 dilution), anti-APC (N terminus; N-15; Santa Cruz Biotech; 1:200 dilution), anti- β -catenin (C19220; Transduction Laboratory; 1:200), anti-PKB (NEB; 1:200 dilution), anti-phospho-PKB (Ser473) (NEB; 1:500 dilution), anti-PCNA (Santa Cruz Biotech; 1:50 dilution), anti-p110 γ (Santa Cruz Biotech; 1:50 dilution). The binding of primary antibodies was

visualized using peroxidase-conjugated goat anti-rabbit or mouse IgG. All sections were counter-stained with haematoxylin. *In situ* hybridization of human colon specimens using human *p110 γ* sense and antisense riboprobes was performed as described²⁴.

Western blotting

Protein extracts from tumours and cell lines were prepared and lysed as described⁸. The following antibodies were used for western blotting: anti-p110 γ (N15, N16, H199), anti-p110 α (H201), anti-p110 β (S19), Cyclin D1 (C20), c-Myc (C8), Smad4 (B8), p53 (FL393), MSH2 (N20), MLH1 (C20), FasL (R20), p21 (C19) (all Santa Cruz Biotech), anti- β -catenin (C19220), Bcl-2 (B46620), Bcl-XL (B22630), Cdk2 (C18520), Cdk4 (C18720), p27 (K25020) (all from Transduction Laboratories). Antibodies against p85 (06-497), Cyclin E (06-459), Cyclin A (06-138) are from Upstate Biotech. Antibodies to SMAD1, SMAD2, and SMAD3 were a gift of C. Sirard, Ontario Cancer Institute.

Transfections, colony formation, cell cycle distribution and *in vivo* tumorigenicity assays

The human HCT116, DLD-1 and HT29 colon cancer cell lines were transfected with a control mammalian expression plasmid, pcDNA3 (Invitrogen), pcDNA3 containing wild-type porcine *p110 γ* , or pcDNA3 containing a kinase-dead (R947P) mutant of porcine *p110 γ* (ref. 10) using Lipofectin (Life Technologies). After 48 h, equal numbers of transfected cells were plated and grown in selection medium (RPMI-1640, 10% FCS, 0.5 mg ml⁻¹ G418) for 2 weeks, after which numbers of colonies were scored. To avoid clonal variations, all the surviving colonies of each transfection were pooled and maintained in the presence of 0.4 mg ml⁻¹ G418. For measurement of DNA synthesis, serum-starved cells were cultured in triplicate in 96-well tissue culture plates (7.5 $\times$ 10³ cells per well) for different time periods and pulsed for the last 8 h with 1 μ Ci per well [³H]thymidine (Amersham). *In vivo* tumorigenicity was analysed by subcutaneous injection of 3 $\times$ 10⁵ cells in 200 μ l of phosphate-buffered saline into both flanks of 4-week-old athymic *nu/nu* mice. Tumours were measured in two dimensions with linear calipers at the indicated days. For cell-cycle distribution, cells were seeded on 6-well tissue-culture plates at 5 $\times$ 10⁵ cells per well, serum-starved for 24 h and then activated with RPMI-1640 containing 5% FCS in duplicate for 24 h. Cells were fixed and stained with propidium iodide to determine cell-cycle phases by flow cytometry.

Received 2 May; accepted 14 June 2000.

1. Tokar, A. & Cantley, L. C. Signalling through the lipid products of phosphoinositide-3-OH kinase. *Nature* **387**, 673–676 (1997).
2. Leever, S. J., Vanhaesebroeck, B. & Waterfield, M. D. Signalling through phosphoinositide 3-kinases: the lipids take centre stage. *Curr. Opin. Cell Biol.* **11**, 219–225 (1999).
3. Cantley, L. C. & Neel, B. G. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc. Natl Acad. Sci. USA* **96**, 4240–4245 (1999).
4. Maehama, T. & Dixon, J. E. PTEN: a tumour suppressor that functions as a phospholipid phosphatase. *Trends Cell Biol.* **9**, 125–128 (1999).
5. Franke, T. F., Kaplan, D. R. & Cantley, L. C. PI3K: downstream AKT1 blocks apoptosis. *Cell* **88**, 435–437 (1997).
6. Shayesteh, L. *et al.* PIK3CA is implicated as an oncogene in ovarian cancer. *Nature Genet.* **21**, 99–102 (1999).
7. Di Cristofano, A., Pesce, B., Cordon-Cardo, C. & Pandolfi, P. P. Pten is essential for embryonic development and tumour suppression. *Nature Genet.* **19**, 348–355 (1998).
8. Sasaki, T. *et al.* Function of PI3K γ in thymocyte development, T cell activation, and neutrophil migration. *Science* **287**, 1040–1046 (2000).
9. Stoyanov, B. *et al.* Cloning and characterization of a G protein-activated human phosphoinositide-3 kinase. *Science* **269**, 690–693 (1995).
10. Stephens, L. R. *et al.* The G β sensitivity of a PI3K is dependent upon a tightly associated adaptor, p101. *Cell* **89**, 105–114 (1997).
11. Li, Z. *et al.* Roles of PLC- β 2 and - β 3 and PI3K γ in chemoattractant-mediated signal transduction. *Science* **287**, 1046–1049 (2000).
12. Hirsch, E. *et al.* Central role for G protein-coupled phosphoinositide 3-kinase γ in inflammation. *Science* **287**, 1049–1053 (2000).
13. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
14. MacPhee, M. *et al.* The secretory phospholipase A2 gene is a candidate for the Mom1 locus, a major modifier of ApcMin-induced intestinal neoplasia. *Cell* **81**, 957–966 (1995).
15. Bernstein, H. G., Keilhoff, G., Reiser, M., Friess, S. & Wetzker, T. Tissue distribution and subcellular localization of a G-protein activated phosphoinositide 3-kinase. An immunohistochemical study. *Cell Mol. Biol.* **44**, 973–983 (1998).
16. White, R. L. Tumor suppressing pathways. *Cell* **92**, 591–592 (1998).
17. Zhou, S., Kinzler, K. W. & Vogelstein, B. Going mad with Smads. *N. Engl. J. Med.* **341**, 1144–1146 (1999).
18. Korinek, V. *et al.* Constitutive transcriptional activation by a β -catenin-Tcf complex in APC^{-/-} colon carcinoma. *Science* **275**, 1784–1787 (1997).
19. He, T. C. *et al.* Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509–1512 (1998).
20. Tetsu, O. & McCormick, F. β -catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
21. Groden, J. *et al.* Response of colon cancer cell lines to the introduction of APC, a colon-specific tumor suppressor gene. *Cancer Res.* **55**, 1531–1539 (1995).
22. Boudreau, M. C., Wilpart, M. & Faivre, J. Diet and colorectal cancer. *Eur. J. Cancer Prev.* **1**, 13–20 (1991).
23. Rudolph, U. *et al.* Ulcerative colitis and adenocarcinoma of the colon in Gai2-deficient mice. *Nature Genet.* **10**, 143–150 (1995).
24. Fata, J. E., Leco, K. J., Moorehead, R. A., Martin, D. C. & Khokha, R. Timp-1 is important for epithelial proliferation and branching morphogenesis during mouse mammary development. *Dev. Biol.* **211**, 238–254 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Saunders for scientific editing and J. Ho, K. Jazier, M. Crackower, A. Oliveira-dos-Santos, L. Zhang, N. Joza, C. Krawczyk, I. Kozieradzki, M. Cheng, R. Sarao, Y.-Y. Kong, M. Nghiem, Q. Liu, E. Griffith, R. Williams, C. Sirard, V. Stambolic, M. Reth, C. Potten, A. Nepren, H. Okada, Y. Jang, S. Pownall, D. Lacey and W. Boyle for reagents and helpful discussions. This work is supported by grants from Amgen, the National Cancer Institute of Canada and the Canadian Center of Excellence for Tumor Vaccination.

Correspondence and requests for materials should be addressed to J.M.P. (e-mail: jpenning@amgen.com).

Accumulation of autophagic vacuoles and cardiomyopathy in LAMP-2-deficient mice

Yoshitaka Tanaka^{*†‡}, Gundula Guhde^{*†}, Anke Suter^{*}, Eeva-Liisa Eskelinen[§], Dieter Hartmann[¶], Renate Lüllmann-Rauch[¶], Paul M. L. Janssen[#], Judith Blanz^{*}, Kurt von Figura^{*} & Paul Saftig^{*}

^{*} Zentrum Biochemie und Molekulare Zellbiologie, Abt. Biochemie II, Universität Göttingen, Heinrich-Düker-Weg 12, 37073 Göttingen, Germany
[§] University of Dundee, Department of Biological Sciences, Dundee DD1 4NH, UK

^{||} Institute of Biotechnology, University of Helsinki, 0001A Helsinki, Finland

[¶] Anatomisches Institut, Christian Albrechts Universität Kiel, 24118 Kiel, Germany

[#] Abteilung Kardiologie und Pneumologie, Universität Göttingen, Robert-Koch-Strasse 40, D-37075 Göttingen, Germany

[†] These authors contributed equally to this work

Lysosome-associated membrane protein-2 (LAMP-2) is a highly glycosylated protein and an important constituent of the lysosomal membrane¹⁻⁷. Here we show that LAMP-2 deficiency in mice increases mortality between 20 and 40 days of age. The surviving mice are fertile and have an almost normal life span. Ultrastructurally, there is extensive accumulation of autophagic vacuoles in many tissues including liver, pancreas, spleen, kidney and skeletal and heart muscle. In hepatocytes, the autophagic degradation of long-lived proteins is severely impaired. Cardiac myocytes are ultrastructurally abnormal and heart contractility is severely reduced. These findings indicate that LAMP-2 is critical for autophagy. This theory is further substantiated by the finding that human LAMP-2 deficiency⁸ causing Danon's disease is associated with the accumulation of autophagic material in striated myocytes.

The collective abundance of LAMP-2 and the structurally and evolutionarily related LAMP-1 has been estimated to be high enough to form a nearly continuous carbohydrate coat on the inner surface of the lysosomal membrane⁷. The functions of LAMP-1 (ref. 9) and LAMP-2 are, however, unknown. The X-chromosomal *lamp-2* gene is ubiquitously expressed^{3,4}. To investigate the function of LAMP-2 we have generated mice deficient for this protein. We constructed a gene-targeting vector to disrupt LAMP-2, and used homologously recombined embryonic stem cell clones to generate chimaeras with germline transmission of the mutated allele (see Supplementary Information). Northern (Fig. 1a) and western

blotting (Fig. 1b) demonstrated a complete absence of *lamp-2* RNA and LAMP-2 protein in homozygous LAMP-2-deficient mice.

Genotyping of offspring from crosses of wild-type males with heterozygote females revealed a frequency of 26% for hemizygote mutant males (LAMP-2^{-/-}), resembling the expected mendelian frequency. These males were bred with heterozygote females to obtain LAMP-2^{-/-} female mice. Offspring of this breeding resulted in 24% hemizygote deficient males, 26% homozygote deficient females, 22% hemizygote wild-type males and 28% heterozygote females. Heterozygotes are fertile and do not show increased mortality (data not shown). LAMP-2-deficient females were fertile

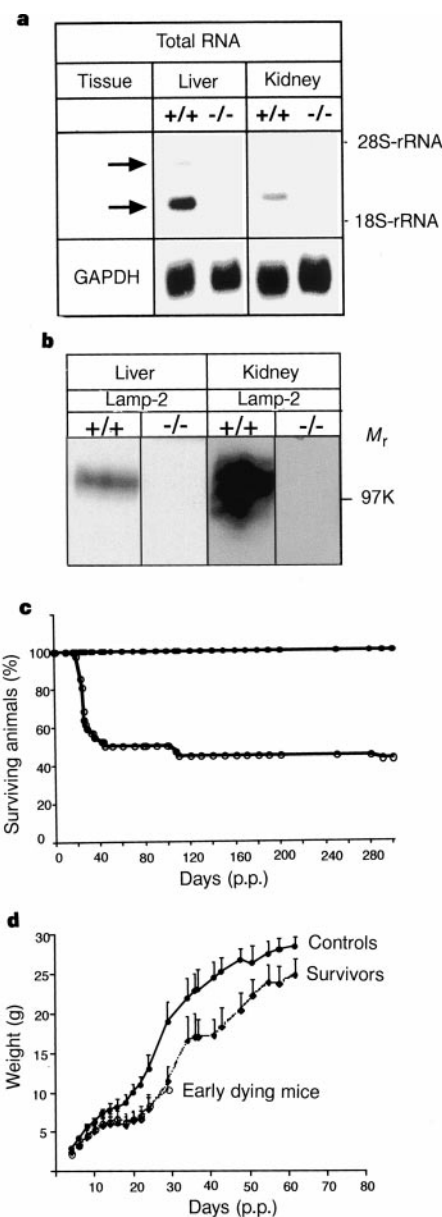


Figure 1 LAMP-2 deficiency causes increased mortality and loss of weight. **a**, Northern blot analysis of *lamp-2* expression. Total RNA was hybridized using a *lamp-2* cDNA² probe and a murine glyceraldehyde-3-phosphate dehydrogenase probe. **b**, Western blot analysis of LAMP-2 expression. In control liver and kidney extracts the glycosylated LAMP-2 molecules were detected. In LAMP-2^{-/-} tissues no LAMP-2 product was found. **c**, Increased mortality in LAMP-2 deficient outbred (C57B6/Jx129SV) mice. Forty-two control (closed circles) and LAMP-2-deficient mice (open circles) were monitored over 300 days. **d**, Loss of weight in early dying ($n = 7$ males) and surviving LAMP-2-deficient mice ($n = 11$ males) in comparison with control mice ($n = 12$ males) monitored for 60 days after birth.

[‡] Present addresses: Kyushu University, Graduate School of Pharmaceutical Sciences, Pharmaceutical Cell Biology, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Japan (Y.T.) and Johns Hopkins University, Institute of Molecular Cardiobiology, Baltimore, Maryland 21205, USA (E.-L.E.).

although the mean litter size was reduced to 4 or 5 animals, compared with 6 or 7 in control mice (not shown). LAMP-2 knockout mice show elevated mortality (Fig. 1f) and reduced weight (Fig. 1c) compared with their wild-type littermates. About 50% of the LAMP-2 deficient animals die between postnatal day 20 and 40, independent of sex and genetic background (C57B6/Jx129SV and 129SV). In addition, LAMP-2-deficient mice are smaller than the wild type. The weight difference is maximal (35–45%) between day 20 and 30. In older mice (over 60 days), the weight difference is 10–15% (Fig. 1d). Mice that die early (4 weeks of age) often have stenoses or segmental haemorrhagic infarction of the small intestine (not shown), which together with pancreatic lesions (see below) could account for their poor weight gain and increased mortality during and after weaning. In lymphoid tissues (thymus and spleen) we observed a pathologically increased apoptotic cell loss (thymus) and defects in the demarcation of white and red pulp (spleen).

The accumulation of autophagic vacuoles is the most prominent finding in liver, kidney, pancreas and cardiac and skeletal muscle in both the mice that die early and the survivors. In 1-, 6- and 19-

month-old LAMP-2-deficient mice, the exocrine pancreas showed the highest degree of autophagic alteration. Most acinar gland cells of the LAMP-2-deficient mice contained numerous autophagic vacuoles of up to 10 μm in diameter (Fig. 2a). Hepatocytes displayed large clusters of smaller autophagic vacuoles containing cytosol, endoplasmic reticulum, sparse glycogen and mitochondria (Fig. 2b). Many of these structures resembled early autophagic vacuoles (Fig. 2c and insert) as defined by the morphologically intact cytoplasmic contents and a double limiting membrane^{10,11}. Autophagic vacuoles were also present in capillary endothelium of kidney, skeletal and cardiac muscles, intestinal wall and lymph nodes and in many neutrophilic leucocytes (not shown).

LAMP-2 deficient skeletal (Fig. 3a) and heart muscle (Fig. 3b) showed an accumulation of vacuoles filled with polymorphic contents. In mice that died early (not shown), the accumulation appeared to be more severe than in the surviving mice. At an age of 19 months, cardiomyocytes were occasionally filled with large vacuoles (Fig. 3b). The skeletal muscle of these mice showed fibre degeneration, fibre splitting and ring fibres (not shown). However, creatine kinase activity in the serum of LAMP-2-deficient mice was in the range of that of controls (not shown).

The *in vitro* findings resemble the *in vivo* findings. Hepatocytes from LAMP-2-deficient mice cultured overnight displayed an accumulation of autophagic vacuoles (Fig. 2c). The vacuoles were classified into early ('Avi'; Figs 2c, insert, and 4b) and late ('Avd', Fig. 4c) autophagic vacuoles, based on morphological criteria^{10,11}. The volume density of early autophagic vacuoles (containing intact cytosol and organelles) was increased about 16-fold, whereas the density of late autophagic vacuoles (containing partially degraded material) was increased about threefold in LAMP-2-deficient hepatocytes (Fig. 4a). Compared with late autophagic vacuoles, early autophagic vacuoles in control cells had a fourfold lower labelling density for cathepsin-D and LAMP-1, markers of the lysosomal matrix and the membrane, respectively (Fig. 4d, e). This difference in labelling density between Avi and Avd was maintained in LAMP-2-deficient hepatocytes, but the labelling intensity for both lysosomal markers was lower than in the control cells (Fig. 4d, e). Autophagosomal pH was not affected by LAMP-2 deficiency. By labelling the hepatocytes with DAMP¹² we estimated that the pH was 6.4 in early and 5.7 in late autophagic vacuoles of control hepatocytes. In LAMP-2-deficient hepatocytes the pH in early autophagic vacuoles was slightly higher (pH 6.7) than in the control cells, and in late autophagic vacuoles it was close (pH 5.8) to that in controls.

Glucagon¹³ and amino-acid deprivation¹⁴ are potent inducers of autophagy. In LAMP-2-deficient mice, the blood serum levels of glucagon and amino acids were within the range of controls except for glutamine, which was increased (7.8 mg dl⁻¹ compared with 5.2 mg dl⁻¹ in controls) and arginine, which was decreased (0.1 mg dl⁻¹ compared with 0.9 mg dl⁻¹ in controls). Although arginine is not one of the amino acids known to induce autophagy, we fed LAMP-2-deficient mice for two weeks with a diet rich in arginine. Serum arginine normalized under the diet, and the accumulation of autophagic vacuoles in liver and pancreas persisted, indicating that autophagy is not induced by low arginine.

We investigated the functional consequences of autophagosome accumulation in hepatocytes and cardiomyocytes. Degradation of long-lived proteins occurs mainly by autophagy¹⁵. In control hepatocytes, long-lived proteins were degraded at a rate of 1.8% h⁻¹. This rate decreased to 0.95% h⁻¹ in LAMP-2-deficient hepatocytes (Fig. 4f). Withdrawal of amino acids and of fetal calf serum increased the degradation rate of long-lived proteins to 3.4% h⁻¹ in control hepatocytes. This increase was completely inhibited when 3-methyladenine, an inhibitor of autophagosome formation¹⁶, was added (Fig. 4g). In LAMP-2-deficient hepatocytes, withdrawal of amino acid and fetal calf serum did not stimulate the degradation of long-lived proteins (Fig. 4g). These data may indicate that the

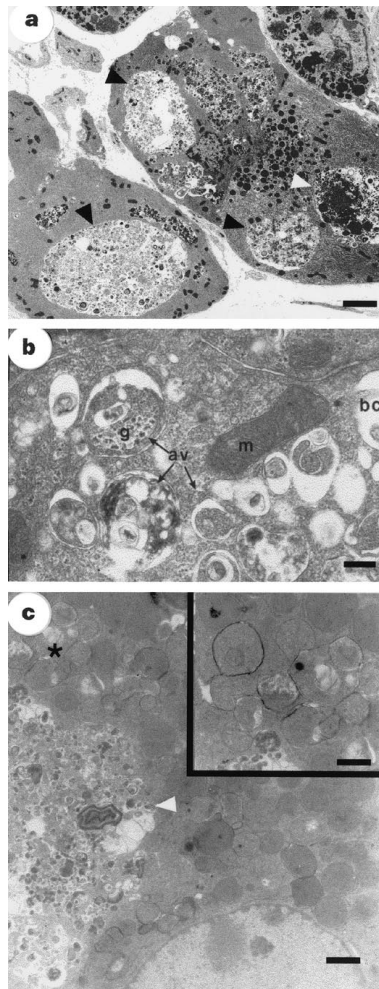


Figure 2 Accumulation of autophagic vacuoles in LAMP-2 deficient tissues. **a**, Ultrastructure of autophagic vacuoles (arrowheads) in pancreatic acinar gland cells from a 6.5-month-old LAMP-2-deficient mouse. **b**, Autophagic vacuoles containing glycogen (g), cytosol and cellular constituents in a hepatocyte of an 8-day-old LAMP-2^{-/-} mouse. m, mitochondria; bc, bile canaliculus. **c**, Accumulation of numerous early autophagic vacuoles in cultured LAMP-2^{-/-} hepatocytes. Arrow, large late autophagic vacuole containing partially degraded material. Insert, magnification of the region indicated by an asterisk with typical early autophagic vacuoles containing undigested cellular constituents. Scale bars: **a**, 3.5 μm ; **b**, 0.2 μm ; **c**, 1.3 μm ; insert, 0.9 μm .

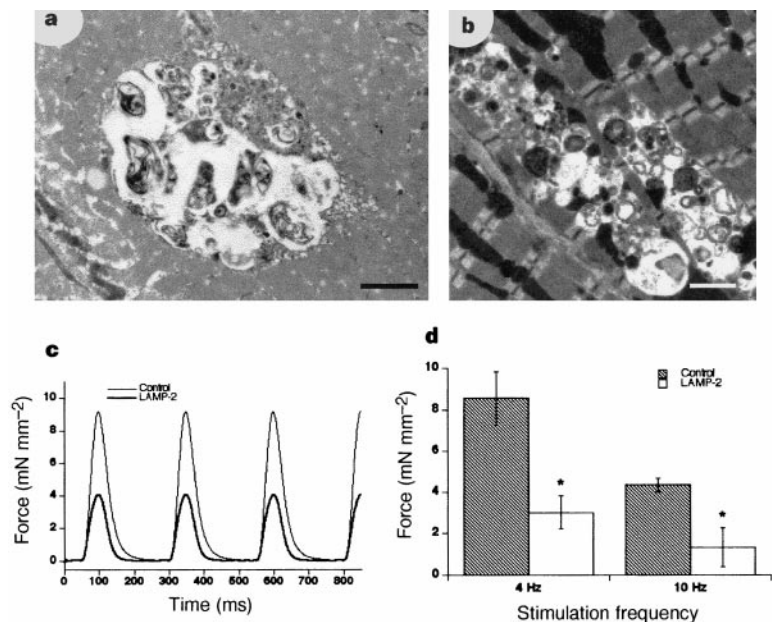


Figure 3 Skeletal and cardiac muscles of a 19-month-old LAMP-2-deficient mouse. **a**, Vacuole in a fibre of the soleus muscle. **b**, Ultrastructure of a vacuole in a cardiomyocyte. Scale bars: **a**, 1 μm ; **b**, 1.5 μm . **c**, Typical original twitch recordings of developed force of heart muscles from control (thin line) and LAMP-2-deficient (thick line)

mice. **d**, Average values of developed force. Preparations of LAMP-2 deficient mice had a significantly lower baseline developed force at a stimulation frequency of 4 Hz and 10 Hz; $P < 0.01$.

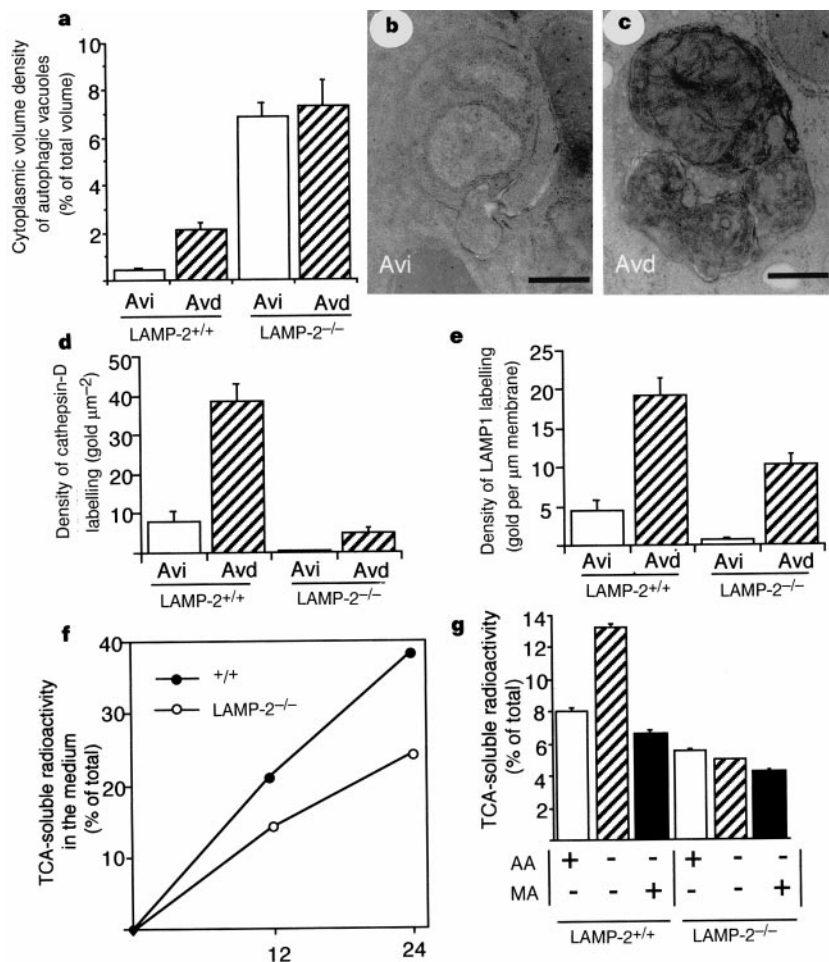


Figure 4 Accumulation of autophagic vacuoles and prolonged half-time of long-lived proteins in LAMP-2-deficient hepatocyte cultures. **a**, Volume density of early (Avi) and late (Avd) autophagic vacuoles from four independent cultures. Examples of Avi and Avd are shown in **b** and **c**, respectively. Bars represent 0.4 μm . **d**, Labelling density of cathepsin-D as estimated in two independent experiments. **e**, Labelling density of LAMP-1.

f, Degradation of long-lived proteins followed over 24 h in amino acid and fetal-calf-serum-containing medium. **g**, Effect of amino-acid (AA) deprivation and 3-methyladenine (MA) on protein degradation during 4 h of chase. The error bars in **a**, **d**, **e** and **g** represent s.e.m.

accumulation of early autophagic vacuoles in LAMP-2-deficient hepatocytes is due to a defect in their maturation to late autophagic vacuoles that actively degrade their content. As a result, stimulation or inhibition of autophagic sequestration of cytoplasmic material has little or no effect upon the degradation rate of long-lived cytoplasmic proteins (Fig. 4g).

LAMP-2-deficient mice had a significantly increased ratio of heart weight (blotted dry) to body weight ($0.51 \pm 0.02\%$ versus $0.82 \pm 0.06\%$, $P < 0.01$), which may be secondary to an insufficient contractile function of heart muscle. The contractile function of heart muscle was severely reduced in LAMP-2-deficient mice (Fig. 3c, d). The contractile developed force at stimulation with 4 Hz was 2.5-fold lower in heart muscle preparations from LAMP-2-deficient mice than in those from controls. At a stimulation frequency of 10 Hz the contractile dysfunction was even more pronounced (Fig. 3d). LAMP-2 deficiency is the primary defect⁸ in the index case of Danon's disease¹⁷ and nine other unrelated patients with this familial disease, which is characterized by fatal cardiomyopathy, variable mental retardation and mild myopathy¹⁸. Ultrastructurally, the accumulation of vacuoles within skeletal and cardiac muscle is a prominent finding. Glycogen accumulates both inside and outside the vacuoles. The presence of cytoplasmic debris within the vacuoles characterizes them as autophagic structures¹⁹. Some aspects of the phenotype of LAMP-2-deficient mice, particularly the accumulation of autophagic vacuoles in striated muscles, resembles that of patients with LAMP-2 deficiency⁸. The phenotypic alterations in LAMP-2 deficient mice clearly go beyond the pathology described so far in the human cases of Danon's disease, in that it encompasses autophagic lesions in non-muscular tissues including neutrophilic leukocytes, hepatocytes and acinar gland cells in pancreas.

In addition to the recent identification of disease-causing mutations in lysosomal membrane proteins leading to nephropathic cystinosis²⁰ and Salla disease²¹, our data point to the vital function of another lysosomal membrane protein. LAMP-2 is required for conversion of early autophagic vacuoles to vacuoles, which rapidly degrade their content. This indicates that LAMP-2 may be involved in the process of fusion of autophagic vacuoles with endosomes/lysosomes, which provide the acid hydrolases required for degradation or a function in the maturation of the autophagolysosomes into actively digesting organelles. How the defect in autophagy relates to the observation that LAMP-2 serves as a receptor for the selective uptake of cytosolic proteins by lysosomes²² remains to be determined. LAMP-2-deficient mice will be a valuable tool to define the functional role of LAMP-2 in autophagy and in selective protein uptake by lysosomes. □

Methods

Generation of LAMP-2-deficient mice

For information describing the generation of the LAMP-2^{-/-} mice, see Supplementary Information or our website (http://www.uni-bc.gwdg.de/bio_2/SAFTIG/Nature2000-Suppl.htm).

Northern and western blot analyses

Total kidney and liver RNA were separated in a formaldehyde agarose gel and processed as described²³. Filters were hybridized with a *lamp-2* complementary DNA² and a 280-base-pair complementary DNA fragment from glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as described²⁴. Western blots for LAMP-2 were done as described⁹ using an anti-mouse LAMP-2 antibody (Abl 93; Developmental Studies Hybridoma Bank).

Hepatocyte primary cultures

Hepatocytes were prepared as described²⁵, enriched using a (58% w/v) Percoll gradient and plated on collagen-coated culture dishes.

Degradation of long-lived proteins

Cultured hepatocytes were incubated in ¹⁴C-valine (0.63 mCi) in valine-free RPMI-1640 medium containing 5% dialysed fetal calf serum (FCS) for 24 h. Cells were washed with Krebs–Heinseleit buffer, and subsequently incubated either in RPMI-1640 with 10% FCS,

or with Krebs–Heinseleit buffer containing 0.1% bovine albumin, and 15 mM cold valine. After 1 h incubation the medium was replaced with the respective fresh medium and the incubation continued for up to 24 h. Cells were washed and suspended in 0.5 ml PBS containing 0.1% Triton X-100. The radiolabelled proteins present in the medium and cells were precipitated with 10% TCA at 0°C. The rate of protein degradation was estimated from the TCA-soluble radioactivity (cells and medium) as percentage of total radioactivity.

Determination of the contractility of the heart muscle

Muscle preparations from 6–9-month-old mice were dissected and mounted in an experimental set-up as described for rat trabeculae²⁶. We measured contractile force under standard conditions at 1.25 mM Ca²⁺, 37°C, and stimulation frequencies of 4 and 10 Hz. Data are expressed as means $\pm$ s.e.m.

Histology

For standard light microscopical histology, immunohistochemistry and transmission electron microscopy tissues were processed as described³. For additional information describing further ultrastructural abnormalities in LAMP-2^{-/-} mice, see Supplementary Information or our website (http://www.uni-bc.gwdg.de/bio_2/SAFTIG/Nature2000-Suppl.htm).

Autophagosome quantification

Cultured hepatocytes were fixed in 2% glutaraldehyde in 0.2 M Hepes, pH 7.4. The cells were scraped off the culture dish and postfixed in 1% OsO₄ for 1 h, dehydrated in ethanol and embedded in Epon. Micrographs (20–25 per sample) were taken with systematic random sampling with a primary magnification of $\times 10,000$. We estimated the cytoplasmic volume fraction of autophagic vacuoles by point counting. Vacuoles were classified as early, containing morphologically intact cytoplasm (Fig. 4b) or late, containing partially degraded but identifiable cytoplasmic material (Fig. 4c)^{10,11}.

Immunogold electron microscopy

Hepatocytes were fixed in 4% paraformaldehyde with or without 0.1% glutaraldehyde in 0.2 M Hepes for 2 h. Cryosections were prepared as described⁹ and labelled with rabbit anti-cathepsin-D or rat anti-Lamp-1 (1D4B, DSHB) antibodies, which were detected with goat anti-rabbit-10 nm gold or goat anti-rat-10 nm gold (British BioCell). To quantify labelling, we took micrographs of autophagic vacuoles (20–25 from each sample) at a primary magnification of $\times 15,000$. The area (cathepsin D) or the length of the limiting membrane (LAMP-1) of the vacuoles was estimated by point counting.

Received 18 February; accepted 23 May 2000.

- Granger, B. L. *et al.* Characterisation and cloning of the Igpn 110, a lysosomal glycoprotein from mouse and rat cells. *J. Biol. Chem.* **265**, 12036–12043 (1990).
- Cha, Y., Holland, S. M. & August, J. T. The cDNA sequence of mouse LAMP-2. Evidence for two classes of lysosomal membrane glycoproteins. *J. Biol. Chem.* **265**, 5008–5013 (1990).
- Konecki, D. S., Foetisch, K., Zimmer, K. P., Schlotter, M. & Lichten-Konecki, U. An alternatively spliced form of the human lysosome-associated membrane protein-2 gene is expressed in a tissue-specific manner. *Biochem. Biophys. Res. Commun.* **215**, 757–767 (1995).
- Aumüller, G., Renneberg, H. & Hasilik, A. Distribution and subcellular localisation of lysosomal-associated protein in human genital organs. *Cell Tissue Res.* **287**, 335–342 (1997).
- Fambrough, D. M., Takeyasu, K., Lippincott-Schwartz, J. & Siegel, N. R. Structure of LEP 100, a glycoprotein that shuttles between lysosomes and the plasma membrane, deduced from the nucleotide sequence of the encoding DNA. *J. Cell. Biol.* **106**, 61–67 (1988).
- Lewis, V. *et al.* Glycoproteins of the lysosomal membrane. *J. Cell Biol.* **100**, 1839–1847 (1985).
- Fukuda, M. Lysosomal membrane glycoproteins. *J. Biol. Chem.* **266**, 21327–21330 (1991).
- Nishino, I. *et al.* Primary LAMP-2 deficiency causes X-linked vacuolar cardiomyopathy and myopathy (Danon's disease). *Nature* **406**, 906–910 (2000).
- Andrejewski, N. *et al.* Normal lysosomal morphology and function in Lysosomal-Associated-Membrane-Protein-Type-1 (LAMP-1) deficient mice. *J. Biol. Chem.* **274**, 12692–12701 (1999).
- Dunn, W. A. Studies on the mechanisms of autophagy: Formation of the autophagic vacuole. *J. Cell Biol.* **110**, 1923–1933 (1990).
- Liou, W., Geuze, H. J., Geelen, M. J. & Slot, J. W. The autophagic and endocytic pathways converge at the nascent autophagic vacuoles. *J. Cell Biol.* **136**, 61–70 (1997).
- Anderson, R. G. Postembedding detection of acidic compartments. *Methods Cell Biol.* **31**, 463–472 (1989).
- Arstila, A. U. & Trump, B. F. Studies on cellular autophagocytosis. The formation of autophagic vacuoles in the liver after glucagon administration. *Am. J. Pathol.* **53**, 687–733 (1968).
- Seglen, P. O. & Gordon, P. B. Amino acid control of autophagic sequestration and protein degradation in isolated rat hepatocytes. *J. Cell Biol.* **99**, 435–444 (1984).
- Henell, F., Berkenstam, A., Ahlberg, J. & Glaumann, H. Degradation of short- and long-lived proteins in perfused liver and in isolated autophagic vacuoles—lysosomes. *Exp. Mol. Pathol.* **46**, 1–14 (1987).
- Seglen, P. O. & Gordon, P. B. 3-Methyladenine: specific inhibitor of autophagic/lysosomal protein degradation in isolated rat hepatocytes. *Proc. Natl Acad. Sci. USA* **79**, 1889–1892 (1982).
- Danon, M. J. *et al.* Lysosomal glycogen storage disease with normal acid maltase. *Neurology* **31**, 51–57 (1981).
- Muntion, F. *et al.* Familial cardiomyopathy, mental retardation and myopathy associated with desmin-type intermediate filaments. *Neuromuscul. Disord.* **4**, 233–241 (1994).
- Morisawa, Y. *et al.* Lysosomal glycogen storage disease with normal acid maltase with early fatal outcome. *J. Neurol.* **160**, 175–179 (1998).
- Town, M. *et al.* A novel gene encoding an integral membrane protein is mutated in nephropathic cystinosis. *Nature Genet.* **18**, 319–324 (1998).

21. Verheijen, F. *et al.* A new gene, encoding an anion transporter, is mutated in sialic acid storage diseases. *Nature Genet.* **23**, 462–465 (1999).
22. Cuervo, A. M. & Dice, J. F. A receptor for the selective uptake and degradation of proteins by lysosomes. *Science* **273**, 501–503 (1996).
23. Köster, A. *et al.* Targeted disruption of the M(r) 46,000 mannose 6-phosphate receptor gene in mice results in misrouting of lysosomal proteins. *EMBO J.* **12**, 5219–5223 (1993).
24. Isbrandt, D. *et al.* Mucopolysaccharidosis VI (Maroteaux-Lamy syndrome): six unique arylsulfatase B gene alleles causing variable disease phenotypes. *Am. J. Hum. Genet.* **54**, 454–463 (1994).
25. Meredith, M. J. Rat hepatocytes prepared without collagenase: Prolonged retention of differentiated characteristics in culture. *Cell Biol. Toxicol.* **4**, 405–425 (1988).
26. Janssen, P. M. L. & Hunter, W. C. Force, not sarcomere length, correlates with prolongation of isosarcometric contraction. *Am. J. Physiol.* **269**, H676–H685 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank N. Leister, A. Wais, M. Grell, G. Jopp and D. Niemeier for technical assistance; K. Rajewski for the E-14-1 cell line; and O. Schunck and K. Nebendahl for veterinary advice. Y.T. was supported in part by the Mochida Memorial Foundation for Medical and Pharmaceutical Research and the Yamanouchi Foundation for Research on Metabolic Disorders. A.S. was supported by a fellowship of the Boehringer Ingelheim Fonds. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to P.S. (e-mail: saftig@uni-bc2.gwdg.de).

Primary LAMP-2 deficiency causes X-linked vacuolar cardiomyopathy and myopathy (Danon disease)

Ichizo Nishino^{*†}, Jin Fu[‡], Kurenai Tanji^{*}, Takeshi Yamada[§], Sadatomo Shimojo^{||}, Tateo Koori[¶], Marina Mora[#], Jack E. Riggs[☆], Shin J. Oh^{**}, Yasutoshi Koga^{††}, Carolyn M. Sue^{*}, Ayaka Yamamoto^{‡‡}, Nobuyuki Murakami^{‡‡}, Sara Shanske^{*}, Edward Byrne^{‡‡}, Eduardo Bonilla^{*}, Ikuya Nonaka[†], Salvatore DiMauro^{*} & Michio Hirano^{*}

Departments of ^{*}Neurology and [‡]Genetics and Development, Columbia University, 630 West 168th Street, P&S 4-443, New York, New York 10032, USA

[†] Department of Ultrastructural Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry (NCNP), 4-1-1 Ogawahigashi-cho, Kodaira, Tokyo 187-8502, Japan

[§] Department of Neurology, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

^{||} Department of Internal Medicine, Saint Marianna University School of Medicine, 2-16-1 Sugao, Miyamae-ku, Kawasaki, Kanagawa 216-8512, Japan

[¶] Department of Pediatrics, Yokohama Rosai Hospital, 321 Kozukue-cho, Kouhoku-ku, Yokohama Kanagawa 222-0036, Japan

[#] Department of Neuromuscular Diseases, National Neurological Institute "C. Besta", Via Celoria 11, Milan 20133, Italy

[☆] Department of Neurology, West Virginia University, PO Box 9180, Morgantown, West Virginia 26506, USA

^{**} Department of Neurology, University of Alabama at Birmingham, UAB Station, Birmingham, Alabama 35294, USA

^{††} Department of Pediatrics and Child Health, Kurume University, 67 Asahi-machi, Kurume, Fukuoka 830-0011, Japan

^{‡‡} Department of Clinical Neuroscience, St. Vincents Hospital, Fitzroy, Victoria 3065, Australia

"Lysosomal glycogen storage disease with normal acid maltase", which was originally described by Danon *et al.*¹, is characterized clinically by cardiomyopathy, myopathy and variable mental retardation. The pathological hallmark of the disease is intracytoplasmic vacuoles containing autophagic material and glycogen in skeletal and cardiac muscle cells. Sarcolemmal proteins and basal lamina are associated with the vacuolar membranes^{2,3}. Here

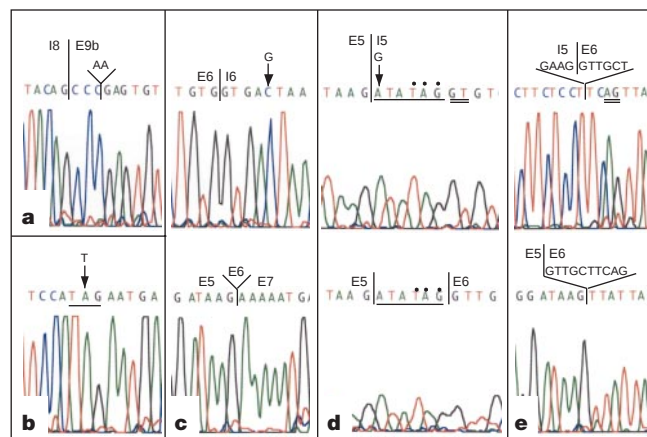


Figure 1 Representative electropherograms of the LAMP-2 gene mutations. **a**, The 2-bp deletion identified in exon 9b of patient 1. **b**, The T-to-A nonsense mutation in exon 4 of patient 2 (stop codon is underlined). **c**, Upper panel, intronic mutation in patient 3; lower panel, sequencing of the RT-PCR product in patient 3, showing exon-6 skipping. **d**, Upper panel, The G-to-A point mutation at the splice-donor site in intron 5 of patient 6; lower panel, sequencing of the RT-PCR product in patient 6, showing a 6-bp insertion at the junction between exons 5 and 6. The inserted sequence (underlined) is derived from the first six nucleotides of intron 5 (underlined), which creates a stop codon (dotted). Most probably, nucleotides 7 and 8, GT (double underlined), in intron 5 are alternatively recognized as a splice-donor site. **e**, Upper panel, the 10-bp deletion in patient 9. This mutation deletes four nucleotides from intron 5 and six from exon 6. Lower panel, sequence of the RT-PCR product in patient 9 reveals the deletion of the 10 nucleotides at the 5' end of exon 6. The 'AG' (double underlined) in exon 6 immediately after the genomic DNA deletion appears to be recognized as an alternative splice-acceptor site.

we report ten unrelated patients, including one of the patients from the original case report¹, who have primary deficiencies of LAMP-2, a principal lysosomal membrane protein. From these results and the finding that LAMP-2-deficient mice manifest a similar vacuolar cardioskeletal myopathy, we conclude that primary LAMP-2 deficiency is the cause of Danon disease⁴. To our knowledge this is the first example of human cardiomyopathy–myopathy that is caused by mutations in a lysosomal structural protein rather than an enzymatic protein.

In 1981, Danon *et al.*¹ reported two boys with the clinical triad of cardiomyopathy, myopathy and mental retardation. Histochemical and electron microscopic features in the boys' muscle mimicked those of Pompe disease but acid maltase activity was normal. Since then, there have been 14 additional case reports in the English literature. Muscle biopsy is diagnostic and shows a characteristic vacuolar myopathy. The vacuoles are limited by a single membrane, contain various 'debris' including cytoplasmic degradation products and glycogen, and stain intensely for acid phosphatase; all are features of lysosomes. However, the vacuolar membrane occasionally merges with indentations of the sarcoplasmic membrane and stains with antibodies to sarcolemmal proteins such as dystrophin and laminin^{2,3}. Moreover, the vacuolar membrane is delineated by basal lamina, which is characteristic of the plasma membrane³. Thus, the vacuoles in this disorder share features of both lysosomes and plasma membrane.

Inheritance of Danon disease (DD) has been considered to be X-linked because in most familial cases males are affected predominantly, affected mothers usually have milder and later-onset cardiac symptoms, and no male-to-male transmission has been described. The X chromosome carries the gene encoding lysosome-associated membrane protein-2 (LAMP-2), which structurally consists of a small cytoplasmic tail with a lysosomal membrane targeting signal⁵, a transmembrane domain, and a large intraluminal head with two internally homologous domains connected by a hinge region rich in proline, serine or threonine—each domain

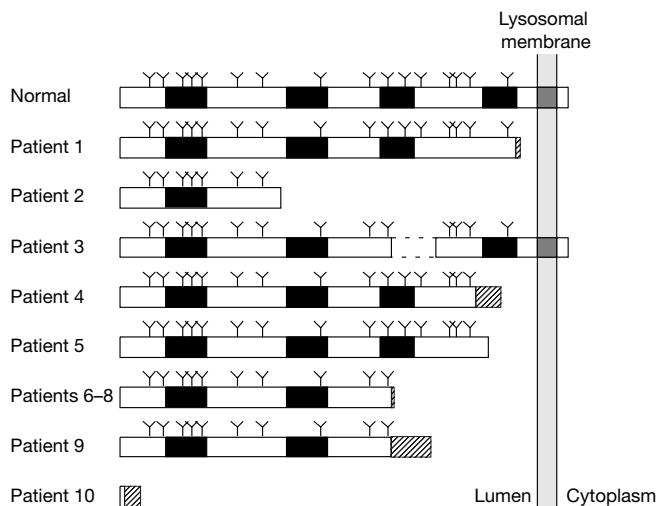


Figure 2 Representation of the mutant LAMP-2b. 'Normal' represents the LAMP-2b amino-acid sequence. Because patient 1, who developed full DD symptoms, has a mutation in exon 9b, the LAMP-2b isoform is most probably responsible for DD. Residue 265, one of the putative disulphide-bond-creating cysteines, and two glycosylation sites are lost by exon-6 skipping in patient 3, which probably results in a significant change in the secondary structure. Black boxes represent the loops created between two cysteine residues flanking each of the four loop regions; Y depicts a glycosylation site; hatched boxes show altered amino acids caused by frameshift mutations; dashes represent missing amino acids due to exon-6 skipping in patient 3. All of the other mutant proteins lack the transmembrane domain (grey bar) because of nonsense or frameshift mutations. Thus, those molecules cannot function as lysosomal 'membrane' proteins.

contains four cysteine residues that form two disulphide bonds. This gene product seemed a particularly promising candidate for the defective protein in DD because 'lysosomes' and 'membranes' are abnormal in this disorder.

The *lamp-2* open reading frame consists of 1,233 nucleotides and encodes 410 amino acids. Exons 1–8 and part of 9 encode a luminal domain; the remainder of exon 9 encodes both a transmembrane domain and a cytoplasmic domain⁵. Human exon 9 exists in two forms, 9a and 9b, which are alternatively spliced and produce two isoforms, LAMP-2a and LAMP-2b, respectively⁶. We sequenced all coding sequences of the LAMP-2 gene in 11 probands and in 54 controls (6 men and 48 women). We also sequenced the exon-flanking sequences in all the samples except for that from patient 10 because of the limited sample availability. We found mutations in all but one proband (Table 1, Fig. 1). We identified eight different mutations: three microdeletions/insertions, two nonsense point mutations, two intronic point mutations and one 10-base-pair (bp) deletion encompassing an exon–intron junction (Fig. 2).

Table 1 Summary of the *lamp-2* mutations

Patient	Ethnic background	Mutation	Site of mutation	Effect on mRNA	Reference
1	Japanese	2-bp del*	E9b	Frame shift	23
2	Japanese	T440A	E4	Non-sense	24
3	Japanese	g5c	I6	E6 skipping	25, 26
4	Italian	T974AA	E8	Frameshift	27
5	African American	C813G	E8	Non-sense	
6	Japanese	g1a	I5	6-bp insertion	
7	African American	g1a	I5	6-bp insertion §	1
8	Anglo-Saxon	g1a	I5	6-bp insertion §	28
9	Japanese	10-bp deletion†	I5/E6 junction	10-bp deletion ‡	
10	Greek	G14del	E1	Frameshift	29

All patients are male and diagnosed as having DD after muscle biopsy. Exonic mutations (upper case) are shown by the nucleotide position in the LAMP-2b cDNA open reading frame. Intronic mutations (lower case) are described by their position in each intron. E, exon; I, intron; del, deletion; and ins, insertion.

* Deletion of AA at nucleotide positions 1,097 and 1,098.

† Deletions of 4 bp at the 3' end of I5 and 6 bp at the 5' end of E6.

‡ Deletion of the sequence corresponding to the 10 nucleotides in the 5' end of exon 6.

§ Not confirmed in these patients.

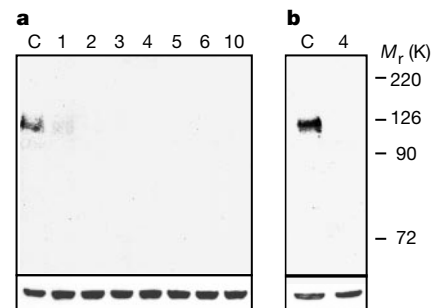


Figure 3 Western blot analysis. **a**, Skeletal muscle. Muscle extracts from patients 1–6 and 10 and a control were blotted and labelled with antibodies against LAMP-2 (top panel) and β -actin (bottom panel). LAMP-2 is deficient in all muscles examined except the sample from patient 1 that showed a trace amount of protein, which is probably the LAMP-2a isoform. In support of this idea, LAMP-2b is the principal LAMP-2 species in skeletal muscle⁶. Numbers denote patients as in Table 1; C, control. **b**, Heart. Protein was extracted from cardiac muscle of patient 4 and a control, blotted onto a nitrocellulose membrane, and labelled with the antibodies against LAMP-2 (top panel) and β -actin (bottom panel). LAMP-2 is deficient also in cardiac muscle. C, control; 4, patient 4.

Both intronic point mutations and the deletion in an exon–intron junction cause aberrant splicing. Although the point mutation in patient 3 does not disrupt the GT–AG rule of splicing, the mutation site in the fifth nucleotide (G) of an intron is highly conserved (86%) in human genes⁷. We carried out polymerase chain reaction with reverse transcription (RT–PCR) and found that exon 6 was skipped in patient 3 (Fig. 1c).

The mutation present in patients 6–8 disrupted the splice donor-site sequence (GT) in intron 5. We sequenced the RT–PCR product from muscle RNA of patient 6 and found a 6-bp insertion at the junction of exons 5 and 6. The inserted nucleotides had the same sequence as the 5' end of intron 5. Conceivably, nucleotides 7 and 8 (GT) immediately after the inserted sequence in intron 5 are recognized as an alternative splice-donor site. The inserted sequence encoded an in-frame stop codon (Fig. 1d), which predicted premature termination of the nascent polypeptide (Fig. 2).

The 10-bp deletion in patient 9 ablated four nucleotides in the 3' end of intron 5, including the splice-acceptor sequence, and six nucleotides in the 5' end of exon 6. This resulted in a 10-bp deletion at the 5' end of exon 6, probably because nucleotides 9 and 10 in exon 6, AG, functioned as a new splice-acceptor site (Fig. 1e). In patients 3, 5, 8 and 9, we also identified a polymorphism (A156T) in exon 2, which does not alter the encoded amino acid.

Although patient 1 had a mutation in exon 9b, which should affect only the LAMP-2b isoform, he had the full syndrome with cardiomyopathy, myopathy and mental retardation. This suggests

that DD is largely due to defects of the LAMP-2b isoform. In support of this idea, LAMP-2b is more abundantly expressed than LAMP-2a in heart, skeletal muscle and brain at the messenger RNA level⁶.

The eight distinct *lamp-2* mutations that we identified, except for the exon-skipping mutation in patient 3, abolish both the transmembrane and the cytoplasmic domains of LAMP-2. The exon-skipping mutation in patient 3 also causes extensive structural changes in LAMP-2, not only because it shortens the protein and ablates two glycosylation sites, but also because it abolishes a loop structure by deleting Cys 265, which is thought to form a disulphide bond with Cys 232 (Fig. 2). All the identified mutations are likely to be equally deleterious because there was no apparent phenotypical differences among our patients and no correlation between the genotypes and phenotypes. We did not find any of these mutations in the DNA sequences of 102 control alleles. The A156T was found in 39 alleles. No other polymorphisms were identified in the open reading frame.

To determine the effects of mutations on protein expression, we carried out western blot analysis, immunohistochemistry, or both, on frozen skeletal muscle specimens from all patients except 7 and 8. Western blot analysis showed complete LAMP-2 deficiency in all muscle specimens studied except for the sample from patient 1. The muscle biopsy had a trace amount of the normal size protein, which was most probably the LAMP-2a isoform because patient 1 had a mutation in exon 9b (Fig. 3a). LAMP-2 was also absent in cardiac muscle from patient 4 (Fig. 3b). By immunohistochemistry, the antibody against LAMP-2 failed to detect any structure in DD muscles, whereas the antibody against limp-I, another lysosomal membrane protein, highlighted the intracytoplasmic vacuoles (Fig. 4). The absence of LAMP-2 immunostaining indicates that the protein is probably not retained in endoplasmic reticulum or other membrane-bound structures.

These results strongly suggest that LAMP-2 deficiency is associated with DD. We found mutations in the LAMP-2 gene in 10 of the 11 patients, a clear indication that at least these 10 individuals have primary LAMP-2 deficiency. One patient, who has LAMP-2 deficiency from their immunohistochemistry without an identified

mutation in the LAMP-2 coding sequence, may have a mutation in non-coding sequences, including the promoter, intron and other untranslated sequences. Alternatively, the LAMP-2 deficiency in this patient might be secondary. The lack of cardiac symptoms in the mother of this patient favours the second explanation.

The LAMP-2 gene is present on the X chromosome in mouse as in human⁸. LAMP-2-deficient mice show an X-linked disease characterized pathologically by numerous autophagic vacuoles⁴. The presence of vacuolar changes similar to DD, albeit not limited to cardiac and skeletal muscle in LAMP-2-deficient mice, supports our conclusion that DD is caused by primary LAMP-2 deficiency in our 10 patients⁴.

The disorders reported¹ as "lysosomal glycogen storage disease with normal acid maltase" are likely to be genetically diverse. Specifically, the rare severely affected infants with unaffected mothers most probably had a different autosomal recessive disease^{9,10}. As glycogen is not always increased in this disease, DD is not a glycogen storage disease. We did not include mental retardation as a defining characteristic because it is not always observed and almost impossible to prove in severely affected patients who die in infancy.

X-linked myopathy with excessive autophagia (XMEA) is a similar vacuolar myopathy characterized by intense deposition of the complement C5b-9 membrane attack complex over the muscle fibre surface. On immunostains, the vacuoles in XMEA are also decorated by antibodies to dystrophin and laminin, suggesting a similar pathomechanism to DD¹¹⁻¹³; however, the vacuoles in an XMEA patient showed strong immunoreactivity both to LAMP-2 and to limp-I, indicating that XMEA is a distinct disease (Fig. 3). In agreement with these findings, sequence analysis of *lamp-2* in another XMEA patient did not show any mutation⁶. Furthermore, the XMEA locus was recently mapped to Xq28 (refs 14, 15), which is distinct from the LAMP-2 locus, Xq24 (ref. 16).

Despite intense investigation, the role of LAMP-2 remains to be fully elucidated. The increased expression of LAMP proteins at the cell surface in metastatic colon cancer and in activated platelets also suggests that it may function in the plasma membrane^{17,18}. In DD,

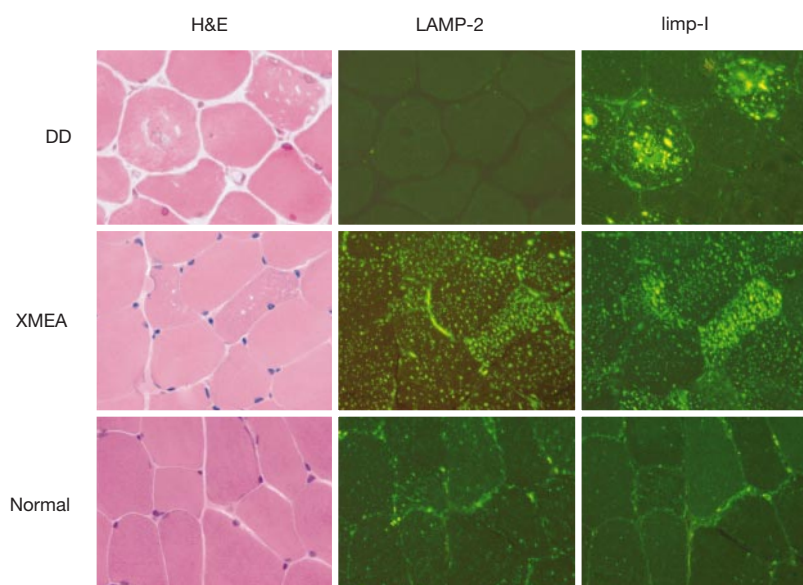


Figure 4 Immunohistochemistry. Serial transverse sections of the biopsied muscle from six DD patients, and samples from a XMEA patient and a normal control, were immunostained with antibodies against LAMP-2 and limp-I. LAMP-2 is absent in all DD muscles (patient 6 shown here), whereas it is present in the XMEA muscle. The antibody against limp-I highlighted vacuoles in both DD and XMEA muscles. LAMP-1 antibody faintly stained both DD and XMEA muscles, suggesting that the expression level of LAMP-

1 might be lower in skeletal muscle than LAMP-2 or limp-I (data not shown). The muscle from the patient without a *lamp-2* mutation also showed LAMP-2 deficiency on immunohistochemistry similar to genetically proven DD patients with primary LAMP-2 deficiency (data not shown). The immunohistochemical stains of LAMP-2, limp-I, and LAMP-1 in normal muscle shows numerous tiny particles and accentuates type 1 fibres. H&E, haematoxylin and eosin, original magnification $\times 150$.

the limiting membrane of the endosomal/lysosomal vacuoles are lined by basal lamina and are often continuous with the plasma membrane; this finding might be due to a defect of normal plasma membrane function. Alternatively, the unusual muscle pathology in DD might result from impaired lysosomal protein processing because the cytoplasmic tail of LAMP-2 functions as a receptor for the uptake of certain proteins into lysosome in association with the relative molecular mass (M_r) 73,000 heat shock cognate protein¹⁹.

Our results, together with the recent identification of mutations in the lysosomal proteins cystinosis and sialin that cause nephropathic cystinosis²⁰ and sialic acid storage diseases²¹, respectively, suggest that lysosomal membrane protein defects might not be rare causes of hereditary human diseases. Although all three diseases manifest prominent intracellular vacuoles, the clinical and detailed histological features are distinct. Nephropathic cystinosis typically begins as a renal tubular Fanconi syndrome in the first year of life progressing to renal failure at 9–10 years of age with prominent intracellular crystal formations representing increased cystine stores²². By contrast, the sialic acid storage disorders, Salla disease and infantile free sialic acid storage disease are predominantly neurodegenerative conditions with single membrane bound vesicles filled with free sialic acid visible as fibrillogranular material²². Sialin and cystinosis are transporter molecules that are thought to export cystine and sialic acid, respectively. By contrast, the function of LAMP-2 is still uncertain. Investigation of this new category of lysosomal diseases will provide important clinical information and yield new insights into the functional roles of the lysosomal membrane and its protein components. □

Methods

Patients

All patients were male and were diagnosed as having DD syndrome on the basis of clinical features and muscle pathology. The four indispensable diagnostic features were cardiomyopathy (confirmed by echocardiogram, electrocardiogram, or both); skeletal muscle weakness on clinical examination; autophagic vacuoles in skeletal muscle; and normal or elevated acid maltase activity. Although not an obligatory feature, mental retardation was seen in eight patients. We excluded one patient with the severe infantile form because inheritance appeared to be autosomal recessive rather than X-linked. Seven of the ten patients were reported previously (Table 1). The patient without *lamp-2* mutation had a total IQ of 92; his mother is healthy, and at age 49 she has a normal electrocardiogram and serum CK level.

Sequence analysis of *lamp-2*

We extracted genomic DNA from blood or skeletal muscle. We amplified each exon and flanking intronic regions with the following primers (the number in the name of the primer indicates exon; F represents forward and R reverse; and sequences are written in 5' to 3' orientation): E1F, GAGATTGGCTGTAAGCAAGA; E1R, GACCAGTCTTT-CAGGTTGTA; E2F, AGTGGTGGGTAGAGCTTGTT; E2R, CAGTCAACGTGGAATTT-CAT; E3F, AAGTGGCATGCCCTACATAA; E3R, GTCAGTGGGAGGGTTATATAA; E4F, GAGAGAAGAAGAAAGCCTAT; E4R, GCCTAGTAGAAGCTATGAAT; E5F, TCTTCCCCTAATATAACCCCTTT; E5R, GGAATCATTTCTAGTAACACTAC; E6F, AGGCTTTGCTGGCCACTTTT; E6R, CTCCTCCCATGCAACTATTT; E7F, GCTTTTCTCTGTGGGTATT; E7R, GGTATGATCAAGGTACACTT; E8F, CCTGGCCAACCTTTCCCATTT; E8R, GAAAGGCCACTGTCAACAT; E9aF, GGAAGTGCTGTCAATTTACT; E9aR, CTCAAAATGCTGGGATGAT; E9bF, CTTTGACTTCGAGACATTCT; E9bR, GTTGACCAAGTATTCATGTT. We directly sequenced the amplified fragments with both forward and reverse primers using BigDye Terminator Cycle Sequencing Kit (PE Biosystems), and then electrophoresed the samples using an ABI PRISM 310 Genetic Analyzer (PE Biosystems). We used PAC clone DJ318C15 sequence (accession number AC002476) as a reference, which includes the entire genomic sequence of LAMP-2 gene in complementary orientation. To determine the effect of the intronic mutations on messenger RNA, we performed RT-PCR in patients 2, 6 and 9. Total RNA was extracted from frozen muscle using Totally RNA Kit (Ambion) and was reverse-transcribed into cDNA with oligo (dT)₂₀ primer using the ThermoScript RT-PCR System (Life Technologies). We amplified the cDNA encompassing from exon 1 to exon 9b using primers E1F, TTTCCCTGGTGTGTCAGCTGTTGTT, and E9bR, GAGTCTGATATCCAGCATAACTTTT. The amplified fragment was directly sequenced using the PCR primers and relevant internal primers. In patient 10, we only sequenced the RT-PCR product from muscle RNA due to the limited sample availability.

Western blotting

We carried out western blot analyses on skeletal muscle from patients 1–6 and 10, and cardiac muscle from patient 4. Western blot analyses were not performed on patients 7–9 because available muscle samples were inadequate. Frozen tissues were briefly washed with

cold PBS, homogenized with triple-detergent lysis buffer (50 mM tris-HCl, pH 8.0, 150 mM NaCl, 0.02% sodium azide, 0.1% SDS, 100 μ g ml⁻¹ phenylmethylsulfonyl fluoride, 1 μ g ml⁻¹ aprotinin, 1% NP-40, sodium deoxycholate), and spun down. The supernatant was collected and stored at –80 °C until use. We separated 5 μ l samples by 7.5% SDS–polyacrylamide gel electrophoresis and electrotransferred the protein onto nitrocellulose membranes at 100 V for 1 h. The membranes were blocked in buffer containing 13.5% fat-free milk, and overlaid with the monoclonal antibody against LAMP-2 (H4B4). The signals were detected using ECL Plus Western Blotting Detection System (Amersham Pharmacia).

Immunohistochemistry

We studied the skeletal muscle from patients 1–3, 5, 6, 9 and the patient in whom we did not find a mutation with immunohistochemical methods. We also analysed the muscles from a normal individual and one patient with XMEA for comparison. Immunohistochemistry was not carried out on patients 4 and 7–9 because available muscle samples were inadequate. Transverse serial sections of 8 μ m thickness were cut from each frozen muscle; one section was stained with haematoxylin and eosin and the others were immunostained with the monoclonal antibodies against LAMP-2 (H4B4), limp-1 (H5C6) and LAMP-1 (H4A3). The sections for immunostaining were fixed with 4% formaldehyde in 0.1 M CaCl₂ (pH 7.2), followed by dehydration in ethanol series, and blocked with 10% fetal bovine serum in PBS. The monoclonal antibodies were applied with 1:50 dilution. The sections were subsequently incubated with biotinylated anti-mouse IgG (Amersham Pharmacia) with 1:100 dilution and fluorescein-conjugated streptavidin (Amersham Pharmacia) with 1:250 dilution. Signals were detected under Optiphot-2 fluorescent microscope (Nikon). All the antibodies used in this study were developed by J. T. August and J. E. K. Hildreth and were obtained from Developmental Studies Hybridoma Bank (University of Iowa).

Received 16 February; accepted 23 May 2000.

- Danon, M. J. *et al.* Lysosomal glycogen storage disease with normal acid maltase. *Neurology* **31**, 51–57 (1981).
- Muntioni, F. *et al.* Familial cardiomyopathy, mental retardation and myopathy associated with desmin-type intermediate filaments. *Neuromusc. Disord.* **4**, 233–241 (1994).
- Murakami, N. *et al.* Sarcolemmal indentation in cardiomyopathy with mental retardation and vacuolar myopathy. *Neuromusc. Disord.* **5**, 149–155 (1995).
- Tanaka, Y. *et al.* Accumulation of autophagic vacuoles and cardiomyopathy in LAMP-2 deficient mice. *Nature* **406**, 902–906 (2000).
- Fukuda, M. Biogenesis of the lysosomal membrane. *Subcell. Biochem.* **22**, 199–230 (1994).
- Konecki, D. S., Foetisch, K., Zimmer, K. P., Schlotter, M. & Lichter-Konecki, U. An alternatively spliced form of the human lysosome-associated membrane protein-2 gene is expressed in a tissue-specific manner. *Biochem. Biophys. Res. Commun.* **215**, 757–767 (1995).
- Lewin, B. *Genes VI* (Oxford Univ. Press, Oxford, 1997).
- Sutton, K. A. & Wilkinson, M. F. The rapidly evolving Pem homeobox gene and Agtr2, Ant2, and Lamp2 are closely linked in the proximal region of the mouse X chromosome. *Genomics* **45**, 447–450 (1997).
- Verloes, A. *et al.* Nosology of lysosomal glycogen storage diseases without *in vitro* acid maltase deficiency. Delineation of a neonatal form. *Am. J. Med. Genet.* **72**, 135–142 (1997).
- Morisawa, Y. *et al.* Lysosomal glycogen storage disease with normal acid maltase with early fatal outcome. *J. Neurol. Sci.* **160**, 175–179 (1998).
- Villanova, M. *et al.* X-linked vacuolated myopathy: complement membrane attack complex on surface membrane of injured muscle fibers. *Ann. Neurol.* **37**, 637–645 (1995).
- Louboutin, J. P. *et al.* Elevated levels of complement components C5 and C9 and decreased antitrypsin activity in the serum of patients with X-linked vacuolated myopathy. *Muscle Nerve* **19**, 1144–1147 (1996).
- Louboutin, J. P., Villanova, M., Lucas-Heron, B. & Fardeau, M. X-linked vacuolated myopathy: membrane attack complex deposition on muscle fiber membranes with calcium accumulation on sarcolemma. *Ann. Neurol.* **41**, 117–120 (1997).
- Villard, L. *et al.* Linkage of X-linked myopathy with excessive autophagy (XMEA) to Xq28. *Eur. J. Hum. Genet.* **8**, 125–129 (2000).
- Auranen, M. *et al.* X-linked vacuolar myopathies: two separate loci and refined genetic mapping. *Ann. Neurol.* **47**, 666–669 (2000).
- Mattei, M. G., Matterson, J., Chen, J. W., Williams, M. A. & Fukuda, M. Two human lysosomal membrane glycoproteins, h-lamp-1 and h-lamp-2, are encoded by genes localized to chromosome 13q34 and chromosome Xq24–25, respectively. *J. Biol. Chem.* **265**, 7548–7551 (1990).
- Kannan, K. *et al.* Cell surface expression of lysosome-associated membrane protein-2 (lamp2) and CD63 as markers of *in vivo* platelet activation in malignancy. *Eur. J. Haematol.* **55**, 145–151 (1995).
- Saitoh, O., Wang, W. C., Lotan, R. & Fukuda, M. Differential glycosylation and cell surface expression of lysosomal membrane glycoproteins in sublines of a human colon cancer exhibiting distinct metastatic potentials. *J. Biol. Chem.* **267**, 5700–5711 (1992).
- Cuervo, A. M. & Dice, J. F. Lysosomes, a meeting point of proteins, chaperones, and proteases. *J. Mol. Med.* **76**, 6–12 (1998).
- Town, M. *et al.* A novel gene encoding an integral membrane protein is mutated in nephropathic cystinosis. *Nature Genet.* **18**, 319–324 (1998).
- Verheijen, F. W. *et al.* A new gene, encoding an anion transporter, is mutated in sialic acid storage disease. *Nature Genet.* **23**, 462–465 (1999).
- Gahl, W. A., Schneider, J. A. & Aula, P. P. In *The Metabolic and Molecular Bases of Inherited Disease* (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 3763–3797 (McGraw-Hill, New York, 1995).
- Matsumoto, S. *et al.* Hepatic involvement in a case of lysosomal glycogen storage disease with normal acid maltase (translation). *Rinsho Shinkeigaku* **39**, 717–721 (1999).
- Itoh, M. *et al.* A patient with lysosomal glycogen storage disease with normal acid maltase (translation). *No To Hattatsu* **25**, 459–464 (1993).
- Katsumi, Y., Tokonami, F., Matsui, M., Aii, H. & Nonaka, I. A case of glycogen storage disease with normal acid maltase accompanied with the abnormal platelet function (translation). *Rinsho Shinkeigaku* **34**, 827–831 (1994).

26. Katsumi, Y. *et al.* Cerebral oxygen and glucose metabolism in glycogen storage disease with normal acid maltase: case report. *J. Neurol. Sci.* **140**, 46–52 (1996).
27. Dworzak, F. *et al.* Lysosomal glycogen storage with normal acid maltase: a familial study with successful heart transplant. *Neuromusc. Disord.* **4**, 243–247 (1994).
28. Riggs, J. E. *et al.* Lysosomal glycogen storage disease without acid maltase deficiency. *Neurology* **33**, 873–877 (1983).
29. Byrne, E. *et al.* Dominantly inherited cardioskeletal myopathy with lysosomal glycogen storage and normal acid maltase levels. *Brain* **109**, 523–536 (1986).

Acknowledgements

We thank patients and their families for their cooperation. We thank K. Murayama, F. Igarashi, S. Tadesse for technical assistance; E. A. Schon, L. Salvati, F. Pallotti, A. Spinazzola, J. Ojaimi, S. Tsujino, N. Minami, M. Ikezawa, Y. Goto, H. Sugita for helpful discussion; and H. Kato, T. Ohi, M. Itoh, Y. Katsumi, T. Wada, X. Dennett, and A. P. Hays for providing samples. M.H. is supported by grants from the National Institutes of Health and the Muscular Dystrophy Association.

Correspondence and requests for materials should be addressed to I.N.
(e-mail: nishino@ncnp.go.jp).

JOINTLESS is a MADS-box gene controlling tomato flower abscission zone development

Long Mao*, Dilara Begum*, Huey-wen Chuang†, Muhammad A. Budiman*, Eugene J. Szymkowiak‡, Erin E. Irish‡ & Rod A. Wing*

* Clemson University Genomics Institute, 100 Jordan Hall, Clemson, South Carolina 29634, USA

‡ Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242, USA

Abscission is a universal and dynamic process in plants whereby organs such as leaves, flowers and fruit are shed, both during normal development, and in response to tissue damage and stress¹. Shedding occurs by separation of cells in anatomically distinct regions of the plant, called abscission zones (AZs). During abscission, the plant hormone ethylene stimulates cells to produce enzymes that degrade the middle lamella between cells in the AZ. The physiology and regulation of abscission at fully developed AZs is well known^{2,3}, but the molecular biology underlying their development is not. Here we report the first isolation of a gene directly involved in the development of a functional plant AZ. Tomato plants with the *jointless* mutation⁴ fail to develop AZs on their pedicels and so abscission of flowers or fruit does not occur normally. We identify *JOINTLESS* as a new MADS-box gene in a distinct phylogenetic clade separate from those functioning in floral organs. We propose that a deletion in *JOINTLESS* accounts for the failure of activation of pedicel AZ development in *jointless* tomato plants.

In tomato, the pedicel AZ, an indented region consisting of a band of anatomically distinct cells, is located in the midpoint of the pedicel⁵. The tomato mutation *jointless* (*j*; ref. 4) completely suppresses the formation of pedicel AZs. In addition, *jointless* affects determinate growth; the inflorescence meristems revert to vegetative growth after forming only one or two flowers, resulting in a 'leafy' inflorescence phenotype^{6–8}. *Jointless* has agronomic value and is widely used in the processing tomato industry. The lack of AZs on *jointless* pedicels yields 'stemless' tomato fruit, which aids mechanical harvesting and prevents physical wounding during transpor-

tation. *Jointless* is a simple recessive mutation and is genetically mapped to chromosome 11 (refs 4, 9).

We previously mapped *jointless* to a 3.0 cM interval between restriction-fragment length polymorphism (RFLP) markers TG523 and RPD158 (Fig. 1a; ref. 10). To isolate *jointless*, we constructed a yeast artificial chromosome (YAC) contig encompassing the locus and showed that two YAC ends, TY159L and TY143R, co-segregate genetically with *jointless*¹¹. *Jointless* is contained within 100 kilobases (kb) of TG523, on the basis of a physical to genetic distance ratio of 86 kb/cM. To isolate candidate genes, we used TG523 to screen a wild-type tomato BAC library¹². We found that a 120-kb bacterial artificial chromosome (BAC) clone (240K4) contains TG523 and the *jointless*-co-segregating YAC end TY159L (Fig. 1a). As the size of 240K4 was much larger than the estimated physical distance from TG523 to TY159L/*jointless* (~86 kb), this BAC probably contained the wild-type *jointless* gene. Therefore we shot-gun sequenced and annotated it (L.M., D.B., M.A.B. and R.A.W., unpublished data). We detected 13 open reading frames (ORFs) from 240K4 and evaluated them by sequence similarity searches in GenBank. One ORF, 240K4.12, located within ~30 kb of TY159L, has significant protein sequence similarity to the conserved domains of MADS-box genes, which are important in plant development^{13–15}. We named this ORF *LeMADS* for *Lycopersicon esculentum* MADS-box gene, and investigated it in detail.

To establish a correlation between *LeMADS* and *jointless*, we used a polymerase chain reaction (PCR) product containing the MADS region as a probe to hybridize DNA from the *jointless* near isogenic lines (NILs) LA3021 (wild type) and LA3023 (*jointless*). All five restriction enzymes tested revealed polymorphism between these two NILs (data not shown), indicating that there could be a deletion in the region. We amplified the putative deletion region by PCR and sequenced it. There is a 939-base-pair (bp) deletion in the *jointless* allele in the 5' region of *LeMADS*, including the first 33 bp of the MADS domain (Fig. 1b). As *jointless* is a spontaneous mutation⁴, the correlation of DNA polymorphism between NILs in a gene co-segregating with the *jointless* phenotype indicated that *LeMADS* was a strong candidate for *JOINTLESS*.

For complementation and antisense suppression experiments, we amplified the *LeMADS* complementary DNA from a wild-type tomato cDNA library by PCR and cloned it in both sense and antisense orientations into the binary vector pBI121 behind the CaMV 35S promoter. 5' and 3' rapid amplification of cDNA end (RACE) experiments showed that the cDNA sequence obtained from the cDNA library was full length at 1,010 bp. We used the sense

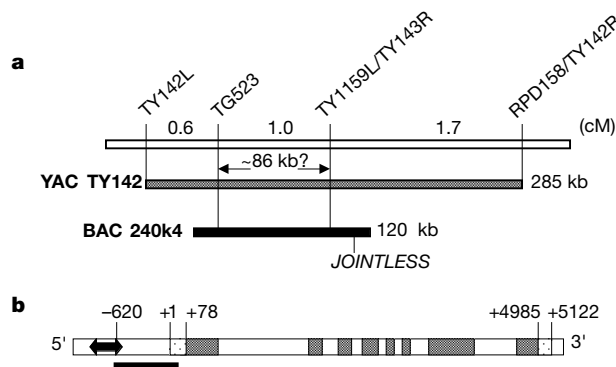


Figure 1 The tomato *JOINTLESS* gene. **a**, Genetic map of the locus on chromosome 11. The TY142L and TY142R are ends of TY142. *Jointless* co-segregates with two other YAC ends TY159L and TY143R. The BAC clone 240k4 contains the flanking genetic marker TG523 and YAC end TY159L. **b**, Diagram of the *JOINTLESS* gene. Black arrow, inverted repeats; +1, transcription start point; dense-dotted boxes, exons; light-dotted boxes, 5' and 3' untranslated regions; solid bar, deleted region in *jointless* allele.

† Current address: Department of Biology, Texas A&M University, College Station, Texas 77843, USA.

construct to transform *jointless* tomato LA3023 for mutant rescue, and tested the antisense construct for its ability to suppress AZ development in wild-type tomato (LA3021).

We obtained 17 sense and 7 antisense transgenic plants. Among the *jointless* tomato plants (LA3023) transformed with the sense construct, the phenotype of 2 out of 17 transgenic plants appeared normal, with AZ-like structures on most flower pedicels (Fig. 2b). After prolonged growth in the greenhouse, eight additional sense transgenic plants developed AZ-like structures on their pedicels. However, the position of restored AZs on the pedicels of transgenic plants varied. In sense transgenic plants, the phenocopied AZs usually developed closer to or even at the base of a flower (Fig. 2b and c, flowers 6–8), rather than at the midpoint of the pedicels as in the wild type. The rescued AZs appeared normal (Fig. 2c, flowers 1, 4, 7 and 8) and were the breakpoints on the pedicels when the flowers died or the fruits ripened. For the remaining sense-transformed plants, we often observed severe bending at the point of presumed AZs (Fig. 2c, flowers 1 and 3). The two plants with nearly perfect restoration of the AZs contained 1–2 copies of the transgene whereas the remaining had more than 2 copies, as shown by Southern analysis. We speculate that the overexpression of *LeMADS* in the transgenic plants with multi-copy sense transgenes caused co-suppression of the expression of *LeMADS*.

Among the wild-type plants (LA3021), transformed with the antisense construct, six out of seven either partially or completely failed to develop AZs on their flower pedicels. We often observed AZ-like bumps in partially suppressed flower pedicels (Fig. 2e). Southern analysis showed that the antisense construct in the seventh plant, whose pedicels looked normal, was rearranged. Unexpectedly, we observed reversions of antisense-suppressed pedicels (Fig. 2f, flowers 1–3) to pedicels with bumps (Fig. 2f, flowers 4 and 6) or AZ-like structures (Fig. 2f, flower 5) in a single inflorescence. The reason for such a reversion to AZ formation is unknown.

To confirm that *LeMADS* was the *JOINTLESS* gene, T1 plants from two antisense and two sense primary transgenic lines were examined. Progeny from both sense and antisense transgenic plants that bore multi-copy transgenes showed *jointless*

phenotypes (co-suppression in the case of sense plants). In contrast, as expected, progeny from the low-copy antisense transgenic T0 plant containing the transgene showed a *jointless* phenotype, whereas those with no transgene, because of genetic segregation, were wild type (Fig. 3a). Similarly, progeny derived from the low-copy sense-rescued primary transgenic plant that inherited the transgene had AZs, but progeny that did not inherit the transgene were *jointless*. The consistent correlation between the *LeMADS* transgene and the *jointless* phenotype shows that *LeMADS* is *JOINTLESS*.

We compared the cDNA sequence of *JOINTLESS* with the genomic sequence from BAC clone 240K4. The genomic DNA of *JOINTLESS* is approximately 6-kb long with eight exons and seven introns (Fig. 1b). There are two inverted repeats in the 5' region that are possibly transposon-like elements responsible for the deletion in the *jointless* allele. This deletion completely eliminated the first exon as the other two potential start codons in this exon are followed by stop codons a few base pairs later. Nevertheless, northern analysis using the 3' sequence of *JOINTLESS* cDNA as a probe shows that gene expression could be detected in *jointless* tomato plants. The major transcript in *jointless* plants is slightly shorter than the wild-type transcript (Fig. 3b). 3' RACE experiments showed no difference between the two transcripts in the 3' end. However, 5' RACE resulted in two bands from both LA3021 and LA3023 (data not shown). Sequencing of the 5' RACE products indicated that the longer band in wild type is part of the normal cDNA. The longer band in *jointless* has a deletion of 115 bp that was replaced with 102 bp of sequence immediately after the deletion point, plus an additional 20 bp reminiscent sequence from the inverted repeat (data not shown). The replacement produced a seven-base larger, aberrant transcript in *jointless* tomato. The smaller 5' RACE products were possibly derived from a secondary transcription start point.

In *jointless* tomato, it appeared that the shorter transcript rather than the aberrant longer transcript accumulated (Fig. 3b). In wild-type tomato, we detected *JOINTLESS* RNA in almost all the tissues. The amount of messenger RNA is highest in shoot tips and axillary buds, where pedicels and inflorescences are developing. However,

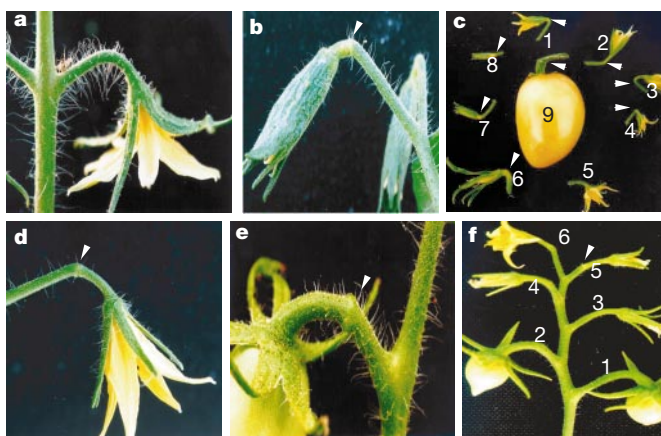


Figure 2 Sense complementation and antisense suppression of AZs in *jointless* and wild-type tomatoes. **a**, A flower from the *jointless* control (LA3023). **b**, A flower with a rescued AZ on its pedicel from a sense construct transformed *jointless* plant. Note that the position of AZ on the pedicel is closer to the flower than those on the wild-type pedicel (Fig. 2d). **c**, Various flowers/fruit from sense-rescued *jointless* plants. Note that flower number 5 is a *jointless* control. **d**, A wild-type flower from jointed control (LA3021). **e**, A pedicel with AZ development partially suppressed by the antisense transgene. The bump-like structure is supposed to be at the position of original AZ in wild-type tomato. **f**, The gradual reversion of suppressed AZs on the same inflorescence from an antisense-transformed wild-type tomato plant from 1 to 6. Arrowheads show the position of the AZ or AZ-like structure.

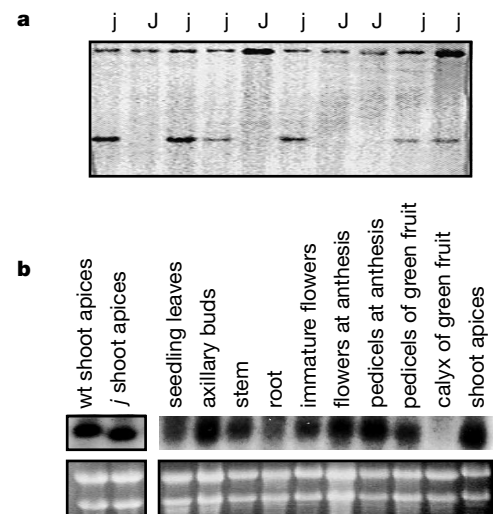


Figure 3 The segregation of transgenes in T1 progeny from an antisense primary transgenic plant (**a**) and expression analysis of *JOINTLESS* gene (**b**). **a**, *EcoRI* and *HindIII* double-digested genomic DNA was probed with *JOINTLESS*MADS domain sequence. The top DNA fragments (about 9.0 kb) are the endogenous genomic DNA containing *JOINTLESS*, and the 1.9 kb fragments below are released cDNA sequence plus the 35S promoter from the transgene. J, wild-type pedicel; j, *jointless*. **b**, mRNA from various tissues of wild-type (wt) tomato and *jointless* tomato shoot tips were hybridized with the 3' sequence of *JOINTLESS*. Note: in the left panel the *jointless* transcript is slightly smaller than that of wild type; tissues in the right panel were all from wild-type tomato.

we also found relatively high amounts of mRNA in fully developed pedicels and flowers (Fig. 3b). The ubiquity of *JOINTLESS* RNA may indicate that the specificity of *JOINTLESS* on pedicel AZ development is regulated at the post-transcriptional level.

A protein sequence search of GenBank revealed that *JOINTLESS* has profound homology with other MADS-box genes. The MADS-box gene most similar to *JOINTLESS* was derived from the *Arabidopsis* Genome Initiative as an annotated ORF from BAC F14M13 (GenBank accession number AC006592). Although there are several MADS-box genes in tomato, they are not among the 20 sequences most similar to *JOINTLESS*, indicating that *JOINTLESS* represents a new kind of MADS-box gene. We developed a phylogenetic tree for *JOINTLESS* with 16 other MADS-box genes from GenBank having the highest matching score (Fig. 4). *JOINTLESS* falls in a distinct group separate from MADS-box genes that are involved in flower development, such as *Arabidopsis* floral homeotic genes *APETALA1* and *CAULIFLOWER*. None of the members inside the clade with *JOINTLESS* are associated with any known phenotypes, though a few, such as *stmads11* and *16* from potato, are expressed mainly in the vegetative tissues¹⁶.

MADS-box genes form a large gene family. They were discovered first in model plants such as *Arabidopsis* and *Antirrhinum*^{13–15}, and homologues have been found in other plants such as *Brassica*¹⁷, rice¹⁸, potato¹⁶ and tomato^{19,20}. Many MADS-box genes are important in the development of plant organs such as floral organs. *JOINTLESS* has a central role in coordinating gene expression underlying the differentiation of the pedicel AZ, and we conclude that it encodes a transcriptional regulator. It is noteworthy that MADS-box genes, such as *AGL8/FUL* (ref. 21) and *SHATTERPROOFs* (ref. 22), are involved in *Arabidopsis* fruit dehiscence zone differentiation. Dehiscence is a process in crucifers with common characteristics to tomato fruit abscission²³. Whether

these genes and *JOINTLESS* have similar targets remains unclear. Furthermore, the *jointless* mutation affects a second important process in plant development, namely maintenance of the inflorescence meristem state. Whether *JOINTLESS* or a second, very closely linked gene has this function is not clear, as LA3021 and LA3023 are also mutant for *self-pruning*, which is epistatic to the leafy inflorescence phenotype of *jointless*⁸. Regardless of the role of *JOINTLESS* in regulating meristem fate, the molecular characterization of *JOINTLESS* will provide important insights into the genetic system that controls the differentiation of an anatomically complex and physiologically important structure, the abscission zone. □

Methods

Sequence analysis

ORFs from BAC 240k4 were searched against GenBank using BlastX²⁴ through the National Center for Biotechnology Information (NCBI) webpage. After cloning the full-length *LeMADS* cDNA, the *LeMADS* protein sequence was deduced. Other MADS-box gene sequences were downloaded from GenBank. Genetics Computer Group (GCG, version 10, Wisconsin Package) program PAUPSEARCH was used for making the phylogenetic tree. Bootstrap analysis using neighbour-joining distance was adopted using the default bootstrap replicates number of 100.

Southern, northern and RACE assays

The MADS conserved region, used as a probe in Southern analysis, was selected according to the genomic sequence and PCR amplified using primers MAOS-1 (5'-CAT TCT CCT CAA TCA TGA CTA A-3') and MAOS-2 (5'-GGT TTA TTC TTT GTT CCC TC-3'). The PCR product was gel purified using a QIAEX II kit (Qiagen) and labelled with ³²P using a DECAprimer II kit (Ambion) for Southern hybridizations. Tomato genomic DNAs were prepared as described²⁵.

Total RNA from LA3021 and LA3023 was prepared from young leaves using RNeasy Plant Mini Kit (Qiagen). 5' RACE was performed following the manufacturer's instructions (5' RACE System, Gibco-BRL, Reagent Assembly v2.0). The three gene-specific primers (GSPs) used were MAOS-CD5 (GSP1), 5'-CTC CCC TCA TTT GCC TTA ATC G-3'; MAOS-CD6 (GSP2), 5'-CTG AAG TTC AAG TGA TGG TTG ATC C-3'; MAOS-CD1r (GSP3), 5'-GCA ACA TCA GCA TCA CAG AGA AC-3'. For 3' RACE, a polyT anchor primer (5'-GGC CAC GCG TCG ACT AGT AC T₍₂₂₎-3') was designed with a similar anchor sequence to that provided with the 5' RACE system. The three GSPs were MAOS-CD1 (GSP1), 5'-GTT CTC TGT GAT GCT GAT GTT GC-3'; MAOS-CD4 (GSP2), 5'-GAA ACA AAT TCT TGA GAG GCG TG-3'; MAOS-CD2r (GSP3), 5'-CA ACA ATT GGA GAG ATC TCT TG-3'.

To avoid cross-hybridization of the MADS-conserved region with other MADS-box genes in tomato, the 3' region of *LeMADS* was amplified using primers MAOS-CD3 (5'-CTC TTG AAA CTG GAT TGA GCC G-3') and MAOS-P3Xba as a probe in northern analysis.

Analysis of the deletion in the *jointless* allele

To determine the approximate region of the deletion in the *jointless* allele, the restriction sites derived from computer analysis of the genomic DNA sequence were compared with Southern data. Nested PCR was performed using primers expected to flank the deletion region. Primers MAOS-3 (5'-TGA TGG ATA TGT ACC TTG AGC-3') and MAOS-4 (5'-GAT AGT TAA AGA TGC GCC TAA C-3'), located about 3.5 kb apart, were used for the first round of PCR. The second round of PCR used primers MAOS-3 and MAOS-5r (5'-GAA TAG ATA TAC ACC CAC ACG-3'). The DNA fragment from LA3023 was cloned into pGEM-T (Promega), and several clones were sequenced on an ABI377 automated DNA sequencer.

cDNA cloning and binary constructs for tomato transformation

To clone a full-length cDNA, primers were designed according to genomic sequence conserved for MADS-box genes and were used together with T7/T3 primers whose sequences were located on the vector used for making the cDNA library, constructed from a wild-type tomato line. PCR reactions were performed directly on phage lysate. By using T3 and MAOS-CD6 (5'-CTG AAG TTC AAG TGA TGG TTG GAT CC-3'), the 78 bp 5' region was obtained from the cDNA library. A fragment amplified with T7 and MAOS-P5 (5'-CCC TCT TTC ATA ACT CTC TTA G-3') from the leading region confirmed all the coding sequences. The PCR products were cloned, sequenced and compared with the genomic sequence.

To make sense and antisense constructs, the cDNA sequence containing all the coding region and most of 5' and 3' untranslated regions was amplified from the plasmid containing the cloned full-length cDNA. Restriction sites were designed at the 5' end of the primers, which were referenced with the restriction sites on the binary vector pBI121. The primers used were MAOS-P5Xba (5'-AGT CTC TAG ACC CTC TTT CTT CAT AC TCT CTT AG-3', underlined is the *Xba*I restriction site), MAOS-P3Sac (5'-GAG TGA GCT CTA CTA ACA ATA ATA TTT ATA TAA AC-3', underlined is the *Sac*I (*Sst*I) restriction site), and MASO-P5Sac and MAOS-P3Xba with the restriction sites swapped. The PCR products of MAOS-P5Xba/P3Sac were digested with *Xba*I and *Sac*I, gel purified and

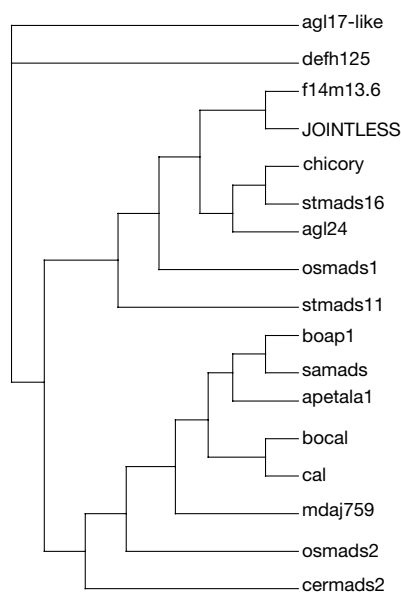


Figure 4 The phylogenetic relationship of *JOINTLESS* with other MADS-box genes. The whole protein sequence of each MADS-box gene was used in the generation of phylogenetic tree. The GenBank accession numbers and brief description of each sequence are: agl17-like, CAB37534, *Arabidopsis* AGL17-like protein; defh125, CAA71739, snapdragon; f14m13.6, C006592_22, *Arabidopsis* BAC F14M13; chicory, AAC84133; stmads16, AAB94005, potato; agl24, AAC63139, *Arabidopsis*; osmads1, BAA81880, rice; stmads11, AAC63139, potato; boap1, Z37968, a cauliflower *apetala-1/squamosa* homologue; samads, Q41276, white mustard; apetala1, *Arabidopsis* AGL7; bocal, AAA64790, *Brassica oleracea* CALIFLOWER; cal, AAA64789, *Arabidopsis* CALIFLOWER; mdmads, CAA04321, apple tree; osmads2, BAA81883, rice; cermads2, BAA25246, fern (*Ceratopteris richardii*).

cloned into *Xba*I, *Sac*I digested binary vector pBI121 to produce sense cDNA constructs. The PCR products of MAOS-P5Sac/P3Xba were made for antisense constructs in pBI121. The constructs were transformed into *Agrobacterium* strain LBA4404 (Gibco BRL) and used later for tomato transformation²⁶.

Copy numbers of transgenes in the transgenic plants were estimated by probing their genomic DNA with labelled NPTII gene sequences amplified with NPTII-F, 5'-ACT GGG CAC AAC AGA CAA TCG-3' and NPTII-R, 5'-AAC GCT ATG TCC TGA TAG CGG-3'. Genomic DNA was digested with *Hind*III and/or *Eco*RI where double digestion was designed to release cDNA from the transgenes.

Received 22 March; accepted 6 June 2000.

- Bleecker, A. B. & Patterson, S. E. Last exit: senescence, abscission, and meristem arrest in *Arabidopsis*. *Plant Cell* **8**, 1169–1179 (1997).
- Addicott, F. T. *Abscission* (Univ. of California Press, Berkeley, Los Angeles, London, 1982).
- Jinn, T. L., Stone, J. M. & Walker, J. C. HAESA, an *Arabidopsis* leucine-rich repeat receptor kinase, controls floral organ abscission. *Genes Dev.* **14**, 108–117 (2000).
- Butler, L. Inherited characters in the tomato. II. Jointless pedicel. *J. Hered.* **37**, 25–26 (1936).
- Roberts, J. A., Schindler, C. B. & Tucker, G. A. Ethylene-promoted tomato flower abscission and the possible involvement of an inhibitor. *Planta* **160**, 159–163 (1984).
- Rick, C. M. & Sawant, A. C. Factor interactions affecting the phenotypic expression of the jointless character in tomatoes. *Am. Soc. Hort. Sci.* **66**, 354–360 (1955).
- Pnueli, L. *et al.* The *SELF-PRUNING* gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. *Development* **125**, 1979–1989 (1998).
- Szymkowiak, E. J. & Irish, E. E. Interactions between *jointless* and wild-type tomato tissues during development of the pedicel abscission zone and the inflorescence meristem. *Plant Cell* **11**, 159–175 (1999).
- Rick, C. M. & Yoder, J. I. Classical and molecular genetics of tomato: highlights and perspectives. *Annu. Rev. Genet.* **22**, 281–300 (1988).
- Wing, R. A., Zhang, H. B. & Tanksley, S. D. Map-based cloning in crop plants. Tomato as a model: I. Genetic and physical mapping of *jointless*. *Mol. Gen. Genet.* **242**, 681–688 (1994).
- Zhang, H. B., Martin, G. B., Tanksley, S. D. & Wing, R. A. Map-based cloning in crop plants. Tomato as a model system II. Isolation and characterization of a set of overlapping yeast artificial chromosomes encompassing *jointless* locus. *Mol. Gen. Genet.* **244**, 613–621 (1994).
- Budiman, M. A., Mao, L., Wood, T. & Wing, R. A. A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Res.* **10**, 129–136 (2000).
- Yanofsky, M. F. *et al.* The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* **346**, 35–39 (1990).
- Sommer, H. *et al.* *Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* **9**, 605–613 (1990).
- Riechmann, J. L. & Meyerowitz, E. M. MADS domain proteins in plant development. *Biol. Chem.* **378**, 1079–1101 (1997).
- Carmona, M. J., Ortega, N. & Garcia-Maroto, F. Isolation and molecular characterization of a new vegetative MADS-box gene from *Solanum tuberosum* L. *Planta* **207**, 181–188 (1998).
- Anthony, R. G., James, P. E. & Jordan, B. R. The cDNA sequence of a cauliflower *apetala-1/squamosa* homolog. *Plant Physiol.* **108**, 441–442 (1995).
- Kang, H. G., Jang, S., Chung, J. E., Cho, Y. G. & An, G. Characterization of two rice MADS box genes that control flowering time. *Mol. Cell* **7**, 559–566 (1997).
- Pnueli, L., Hareven, D., Rounsley, S. D., Yanofsky, M. F. & Lifschitz, E. Isolation of the tomato *AGAMOUS* gene TAG1 and analysis of its homeotic role in transgenic plants. *Plant Cell* **6**, 163–173 (1994).
- Pnueli, L. *et al.* The MADS box gene family in tomato: temporal expression during floral development, conserved secondary structures and homology with homeotic genes from *Antirrhinum* and *Arabidopsis*. *Plant J.* **1**, 255–266 (1991).
- Gu, Q., Ferrandiz, C., Yanofsky, M. F. & Martienssen, R. The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* **125**, 509–517 (1998).
- Liljegren, S. J. *et al.* *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**, 766–770 (2000).
- Petersen, M. *et al.* Isolation and characterisation of a pod dehiscence zone-specific polygalacturonase from *Brassica napus*. *Plant Mol. Biol.* **31**, 517–527 (1996).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Bernatzky, R. & Tanksley, S. D. Methods for detection of single or low copy sequences in tomato on Southern blots. *Plant Mol. Bio. Rep.* **4**, 37–41 (1986).
- McCormick, S. *et al.* Leaf disc transformation of cultivated tomato (*L. esculentum*) using *Agrobacterium tumefaciens*. *Plant Cell Reports* **5**, 81–84 (1986).

Acknowledgements

We thank H.-B. Zhang, S. D. Choi and C. Scheuring for their work during the early stage of this project; C. M. Rick for the NILs LA3023 and LA3021; S. S. Woo for sequence analysis of YAC end TY159L; and the staff of the CUGI DNA Sequencing, Bioinformatics (M. Sasinowsky) and BAC/EST Resource (D. Frisch) Centers. This work is supported by a National Science Foundation grant to R.A.W. and E.J.S. and the Coker Chair in Plant Molecular Genetics to R.A.W. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Correspondence and requests for materials should be addressed to R.A.W. (e-mail: rtwing@clmson.edu). The sequence of *JOINTLESS* is deposited in GenBank under accession number AF275345.

A mutant of the motor protein kinesin that moves in both directions on microtubules

Sharyn A. Endow* & Hideo Higuchi†

* Department of Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA

† Department of Metallurgy and Center for Interdisciplinary Research, Tohoku University, Sendai 980-8579, Japan

Molecular motors move directionally to either the plus or the minus end of microtubules or actin filaments. Kinesin moves towards microtubule plus ends, whereas the kinesin-related Ncd motor moves to the minus ends. The 'neck'—the region between the stalk and motor domain—is required for Ncd to move to microtubule minus ends^{1,2}, but the mechanism underlying directional motor movement is not understood. Here we show that a single amino-acid change in the Ncd neck causes the motor to reverse directions and move with wild-type velocities towards the plus or minus end; thus, the neck is functional but directionality is defective. Mutation of a motor-core residue that touches the neck residue in crystal structures^{2,3} also results in movement in both directions, indicating that directed movement to the minus end requires interactions of the neck and motor core. Low-density laser-trap assays show that a conformational change or working stroke of the Ncd motor is directional and biased towards the minus end, whereas that of the neck mutant occurs in either direction. We conclude that the directional bias of the working stroke is dependent on neck/motor core interactions. Absence of these interactions removes directional constraints and permits movement in either direction.

The Ncd neck mutant that we analysed was based on a *kar3* mutant, *kar3-894*, recovered in a yeast suppressor screen⁴. *kar3-894* has a missense mutation, N378K, in a neck residue that is highly conserved in the carboxy-terminal motor kinesins⁵. We made the corresponding mutation, N340K, in Ncd, and expressed the new

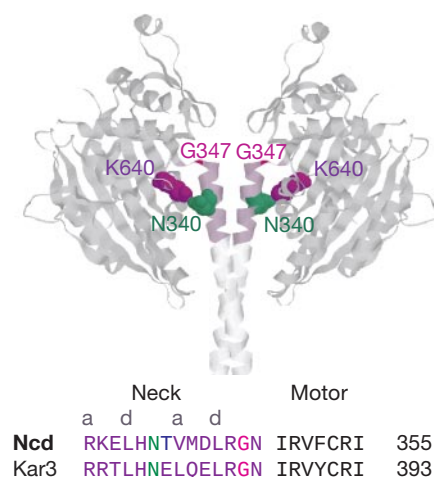


Figure 1 Neck and motor-core mutants. Top, the N340 neck and K640 motor-core residues in a crystal structure² of Ncd. The neck residues^{1,2}, R335–N348 (pink), form a continuous coiled coil with the stalk (white). G347 is near the junction of the neck and motor (grey). The mutants that we analysed are NK11 (N340K), Sup11 (K640N) and NK11 Sup11 (N340K K640N). Bottom, the mutated Ncd neck residue, N340 (green), is conserved in Kar3 and other C-terminal motor kinesins⁵. Residues in the 'a' and 'd' positions (or 1 and 4) of the neck heptad repeat of hydrophobic amino acids are indicated. The motor residues correspond to the first structural element, β 1, of the conserved motor core.

mutant, NK11 (Fig. 1), in bacteria and tested it in microtubule gliding assays.

Motility assays of NK11, in which the motor was bound to the coverslip, showed active gliding of asymmetric axoneme–microtubules consisting of single microtubules grown from the plus end of doublet axoneme fragments. Unexpectedly, about half of the complexes moved with the microtubule leading ($n = 45$ (56%), total = 81), indicating minus-end motor movement, and the rest moved with the axoneme leading ($n = 36$ (44%), total = 81), indicating plus-end movement. The plus- and minus-end movement did not differ significantly from random directionality in χ^2 tests for significance. The single residue change in the Ncd neck thus caused the mutant motor to move randomly towards microtubule plus or minus ends. Five complexes reversed direction one to seven times after gliding in one direction or the other (Fig. 2a). Three complexes broke within the microtubule or at the axoneme–microtubule junction, and the broken halves of one complex glided in opposite directions (Fig. 2b).

Tracking records for the gliding assays showed that the reversals in direction occurred abruptly (Fig. 3). The gliding velocities varied, but the mean velocities were $11.6 \pm 1.3 \mu\text{m min}^{-1}$ (mean $\pm$ s.e.m., $n = 17$) and $14.3 \pm 1.2 \mu\text{m min}^{-1}$ ($n = 11$) for plus- and minus-end movement, respectively, and the peaks of absolute velocity values overlapped that for wild-type Ncd minus-end movement ($16.4 \pm 0.6 \mu\text{m min}^{-1}$ ($n = 23$)). Some axoneme–microtubules glided more than $20 \mu\text{m}$ on the NK11 motor with either the axoneme or microtubule leading. The long excursions of the complexes in one direction or the other are not characteristic of

diffusional movement. The long runs and abrupt reversals could be due to cooperativity of the motors on the glass surface. The breakage was attributed to differing forces on the complex caused by motors moving with different velocities or in opposite directions.

The residue mutated in NK11, N340, touches a motor-core residue, K640, in crystal structures^{2,3} of Ncd (Fig. 1). To test whether the random directionality of NK11 is caused by defective interactions between N340 and K640, we mutated K640 to asparagine. Motility assays of the new mutant, Sup11, showed movement to the minus end of many bound complexes ($n = 17$ (65%), total = 26) but to the plus end of others ($n = 9$ (35%), total = 26). Four complexes reversed directions one or more times, moving first in one direction, and then in the other. The K640N motor-core mutation, like the N340K neck mutation, thus caused the motor to move in either direction on microtubules, indicating that interactions between the two residues are needed to direct Ncd movement to the minus end.

We also analysed the NK11 Sup11 double mutant, N340K K640N, which has the neck and motor-core residues interchanged. Motility assays showed movement to either the microtubule plus ($n = 28$ (38%), total = 73) or minus end ($n = 45$ (62%), total = 73). Ten complexes reversed direction and three broke, as for NK11. The failure of the double mutant to revert the motor to minus-end directionality indicates that the two residues do not interact normally in an inverted configuration.

To determine the basis of the bidirectional movement of the mutant, we carried out assays at low motor density in a laser trap with the Ncd or NK11 motor attached to beads by a biotin–avidin linkage⁶ (Fig. 4). The number of active motors attached to a bead was estimated from the fraction of beads that bound to a microtubule in the presence of ATP and AMP-PNP, assuming random distribution of the motors on the bead. A fraction of 0.50 or less was taken to indicate a probability of 0.02 or less that more than one motor could simultaneously interact with a microtubule⁷. For the

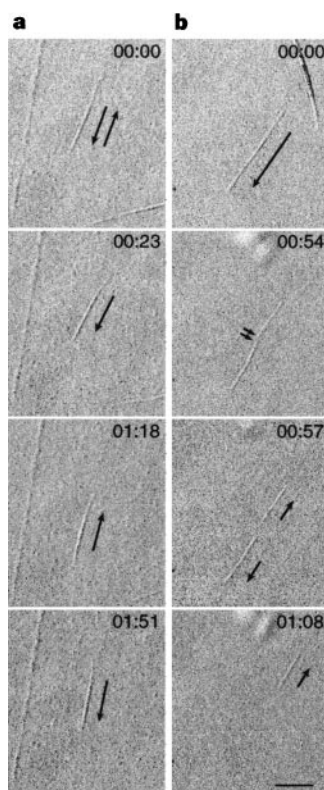


Figure 2 Reversal and breakage of axoneme–microtubules in NK11 gliding assays. **a**, The axoneme–microtubule moved first with the denser axoneme fragment leading and then with the microtubule leading (arrows), and continued to reverse direction, moving in the directions indicated by the arrows. **b**, The complex glided with the axoneme leading (first frame), reversed direction seven or more times, paused and broke (second frame, small arrows), and the two fragments glided in opposite directions (third frame). The axoneme detached from the coverslip and the microtubule continued to glide with the plus end leading (fourth frame). Time shown as minutes:seconds. Scale bar, $5 \mu\text{m}$.

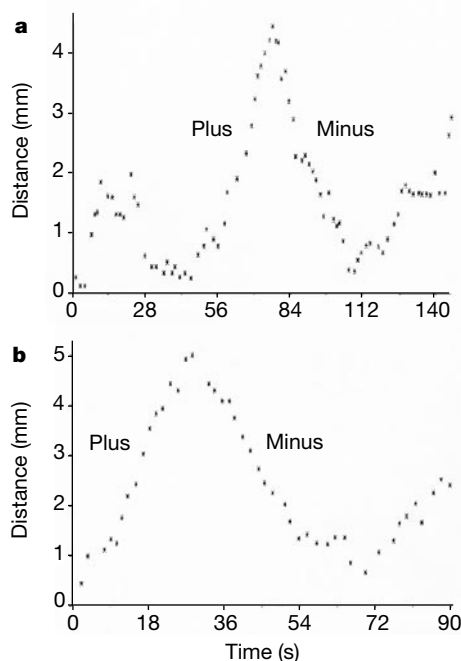


Figure 3 Velocity of gliding on the NK11 mutant motor. **a**, Tracking¹⁸ record for the axoneme–microtubule shown in Fig. 2a. The gliding velocity before and after the main reversal of direction was the same ($6.8 \mu\text{m min}^{-1}$) in the plus (axoneme leading) and minus (microtubule leading) directions. **b**, The velocity of the complex before and after the main reversal of direction was $11.2 \mu\text{m min}^{-1}$ in the plus direction and $8.8 \mu\text{m min}^{-1}$ in the minus direction.

data shown in Fig. 4, the fraction of beads with attached Ncd or NK11 that bound to a microtubule in 10 μ M ATP (Ncd = 0.43, n = 10, total = 23; NK11 = 0.38, n = 9, total = 24) was similar to the fraction in 1 mM AMP-PNP (Ncd = 0.50, n = 8, total = 16; NK11 = 0.47, n = 7, total = 15), a condition under which Ncd binds tightly to microtubules⁸, indicating that almost all of the motors that bind can generate force. Assays in 1 mM ATP showed lower fractions of beads bound to the microtubule (Ncd = 0.13, n = 1, total = 8; NK11 = 0.15, n = 2, total = 13). The data indicated a probability of 98% or greater that single motors were interacting with the microtubule under our assay conditions of 10 μ M ATP.

The displacements that we observed occurred as single events,

indicating that Ncd and NK11 are non-processive motors. Assays at higher motor density showed processive movement. Ncd moved processively towards the minus end, but NK11 moved processively in either direction and a single bead could reverse direction. The difference in motility for beads with low and higher motor densities provides evidence that the displacements in our low-density assays result from single motors.

Traces from the low-density assays of Ncd and NK11 (Fig. 4a) were averaged to reduce brownian noise and determine the times of motor attachment and size of the displacements. The variance of displacements for Ncd showed a shift in distribution towards the microtubule minus end that could be fit with a single gaussian

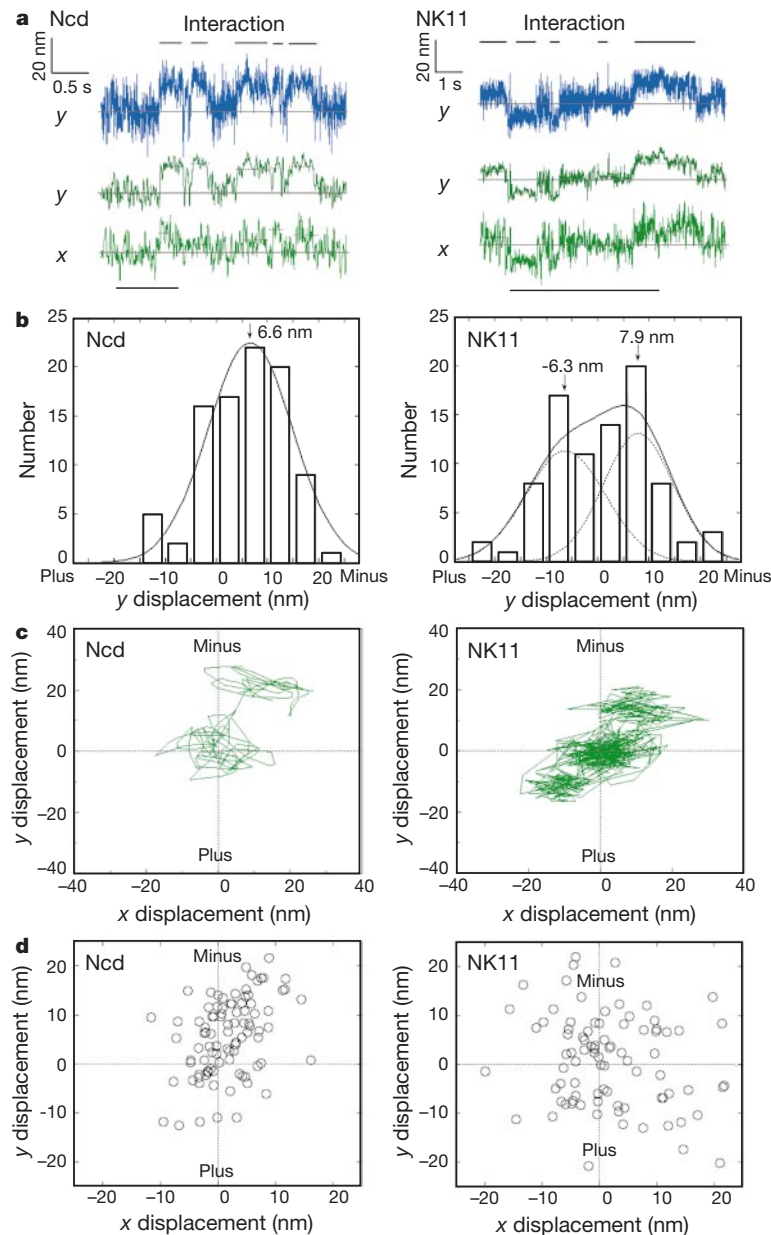


Figure 4 Low-density laser-trap assays of wild-type and mutant motors. **a**, Raw (blue) and filtered (green) traces of Ncd or NK11 interaction with a fluorescent axoneme-microtubule that was positioned vertically with the minus end at the top. Traces show the bead position (nm) in the y and x axes versus time (s). Intervals of motor interaction with the microtubule are shown at the top (bold lines) and in the traces (broken lines), and are correlated with reduced brownian noise. The middle and lower traces were passed through a 50-Hz low-frequency filter. Displacements above and below the y baseline are towards the minus and plus end, respectively. Solid lines below the x traces show the region from which the x-y plots in **c** were made. **b**, Variance of y-axis displacements for

Ncd and NK11. The histogram of displacements is shifted towards the minus end for Ncd, but shows displacements towards both the plus and minus ends for NK11. **c**, x-y plots of motor interaction with a microtubule. The plot for Ncd shows a single displacement towards the microtubule minus end, off-axis to the right. The plot for NK11 shows displacements towards the plus end, off-axis to the left, followed by displacements towards the minus end, one of which is off-axis to the right. **d**, Direction of Ncd and NK11 displacements. The difference in the centre of bead position before and after each movement was determined from x-y plots such as those shown in **c**.

curve, but the variance of displacements for NK11 showed a broadened distribution that was best fit with a double gaussian curve (Fig. 4b). Ncd showed an average displacement of 6.6 ± 0.8 nm to the minus end (mean $\pm$ s.e.m., total = 98), whereas NK11 showed average displacements of 6.3 ± 1.2 nm to the plus end and 7.9 ± 1.1 nm to the minus end (total = 86). The rigidity of the linkage of the motors to the bead was calculated from the amplitude of the brownian noise and found to be stiff, $0.05\text{--}0.10$ pN nm⁻¹, even at zero displacement (Fig. 4a, NK11). With a stiff linkage of the motor to the bead, any change in angle or movement of the motor will be amplified by the displacement of the bead, which acts as a lever arm. We therefore believe that we are measuring a conformational change within the motor, rather than a step of the motor along the microtubule.

x - y plots of the Ncd displacements showed that the movement was directional and biased towards the minus end and to the right (x displacement = 2.5 ± 0.6 nm at $y > 0$ nm, mean $\pm$ s.e.m., $n = 69$), whereas NK11 movement was towards the plus or minus end and frequently to the left or right (Fig. 4c, d). Thus, Ncd attaches and/or makes a conformational change that is biased in one direction, whereas the mutant does so in either direction. Movement of Ncd towards the minus end and a protofilament to the right is consistent with the microtubule rotation caused by Ncd in gliding assays and its handedness⁹. In contrast, NK11 moves towards the plus or minus end and a protofilament to the left or right.

The residues mutated in NK11 and Sup11 are highly conserved among the C-terminal motor kinesins, which may all be minus-end motors¹⁰. The wild-type velocities of the mutant motor indicate that the neck is functional, but directionality is defective. We have shown here that single mutant motors can move on microtubules with either polarity of movement. This finding excludes differences in the motor-tubulin interface as the basis of Ncd 'reversed' directionality. Instead, a working stroke of the motor has a directional bias that displaces the motor towards the minus end. The working stroke of the NK11 neck mutant occurs in either direction, displacing the motor towards either the plus or minus end. The mutants indicate that the directional bias is dependent on neck/motor-core interactions. The directional bias of the wild-type motor could result from a levered movement that is dependent on interactions of the neck residue with the motor-core residue. Interactions that involve residues of the kinesin neck linker¹¹ or structural elements of myosin analogous to the Ncd neck (for example, the converter¹²) could explain the directional movement of kinesin to microtubule plus ends or the reversed movement of myosin VI to actin minus ends¹³. □

Methods

Plasmids

Plasmids for mutant motors were constructed from *pGEX/MC1* (ref. 14) using PCR to introduce mutations and were confirmed by DNA sequencing. Mutants contained glutathione *S*-transferase (GST) fused to the N terminus of Ncd residues 209–700. *pBioNcd* and *pBioNK11* were made by ligating a fragment from the *Propionibacterium shermanii* biotin-containing transcarboxylase gene to Ncd or NK11 DNA. The *pMW172*-based plasmids¹⁵ encode a 120-residue peptide that is biotinylated in bacteria fused to the N terminus of Ncd or NK11 residues 293–700.

Protein expression

We prepared lysates of GST fusion proteins for microtubule gliding assays¹⁶ and analysed them by immunoblotting with an anti-Ncd C-terminus antibody¹⁷. A band of the predicted relative molecular mass for the motor was detected in lysates of all three mutants and bands attributable to breakdown products were absent. BioNcd and BioNK11 proteins for laser-trap experiments were purified by S-Sepharose chromatography¹⁶ followed by microtubule affinity purification. Reaction with Ncd antibody and streptavidin-alkaline phosphatase showed that the proteins were positive for both Ncd and the biotinylated tag. A control Ncd motor protein showed no staining with streptavidin-alkaline phosphatase.

Motility assays

Gliding assays of GST fusion proteins bound to coverslips were in 10 mM NaPO₄ pH 7.4, 1 mM EGTA, 1 mM MgCl₂ and 5 mM Mg-ATP using axoneme-microtubule complexes¹⁸.

Laser-trap assays were carried out by diluting and attaching BioNcd or BioNK11 to beads⁶ in buffer plus 200 mM NaCl, and by monitoring their binding and movement on fluorescent microtubules or asymmetrically labelled fluorescent axoneme-microtubules. Beads were trapped at a trap stiffness of 0.042 pN nm⁻¹ and held next to a microtubule bound to the coverslip surface. Attenuated brownian motion of the bead indicated attachment of a motor on the bead to the microtubule. The standard deviation of brownian noise of the unbound beads was ± 9.9 nm.

Received 31 May; accepted 20 June 2000.

- Endow, S. A. & Waligora, K. W. Determinants of kinesin motor polarity. *Science* **281**, 1200–1202 (1998).
- Sablin, E. P. *et al.* Direction determination in the minus-end-directed kinesin motor ncd. *Nature* **395**, 813–816 (1998).
- Kozlowski, F., De Bonis, S., Burmeister, W. P., Cohen-Addad, C. & Wade, R. H. The crystal structure of the minus-end-directed microtubule motor protein ncd reveals variable dimer conformations. *Structure* **7**, 1407–1416 (1999).
- Hoyt, M. A., He, L., Totis, L. & Saunders, W. S. Loss of function of *Saccharomyces cerevisiae* kinesin-related *CIN8* and *KIP1* is suppressed by *KAR3* motor domain mutations. *Genetics* **135**, 35–44 (1993).
- Endow, S. A. Determinants of molecular motor directionality. *Nature Cell Biol.* **1**, 163–167 (1999). NB: See online version at <http://www.nature.com/ncb/> for correct sequence alignment.
- Iwatani, S., Iwane, A. H., Higuchi, H., Ishii, Y. & Yanagida, T. Mechanical and chemical properties of cysteine-modified kinesin molecules. *Biochemistry* **38**, 10318–10323 (1999).
- Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77**, 773–784 (1994).
- Crevel, I. M., T. C., Lockhart, A. & Cross, R. A. Weak and strong states of kinesin and ncd. *J. Mol. Biol.* **257**, 66–76 (1996).
- Walker, R. A., Salmon, E. D. & Endow, S. A. The *Drosophila* claret segregation protein is a minus-end directed motor molecule. *Nature* **347**, 780–782 (1990).
- Endow, S. A. in *Guidebook to the Cytoskeletal and Motor Proteins* (eds Kreis, T. & Vale, R.) 403–408 (Oxford Univ. Press, Oxford, 1999).
- Kozlowski, F. *et al.* The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. *Cell* **91**, 985–994 (1997).
- Houdusse, A., Kalabokis, V. N., Himmel, D., Szent-Györgyi, A. G. & Cohen, C. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: a novel conformation of the myosin head. *Cell* **14**, 459–470 (1999).
- Wells, A. L. *et al.* Myosin VI is an actin-based motor that moves backwards. *Nature* **400**, 505–508 (1999).
- Chandra, R., Salmon, E. D., Erickson, H. P., Lockhart, A. & Endow, S. A. Structural and functional domains of the *Drosophila* ncd microtubule motor protein. *J. Biol. Chem.* **268**, 9005–9013 (1993).
- Way, M., Pope, P., Gooch, J., Hawkins, M. & Weeds, A. G. Identification of a region in segment 1 of gelsolin critical for actin binding. *EMBO J.* **9**, 4103–4109 (1990).
- Chandra, R. & Endow, S. A. Expression of microtubule motor proteins in bacteria for characterization in *in vitro* motility assays. *Methods Cell Biol.* **39**, 115–127 (1993).
- Hatsumi, M. & Endow, S. A. The *Drosophila* ncd microtubule motor protein is spindle-associated in meiotic and mitotic cells. *J. Cell Sci.* **103**, 1013–1020 (1992).
- Song, H., Golovkin, M., Reddy, A. S. N. & Endow, S. A. *In vitro* motility of AtKCBP, a calmodulin-binding kinesin protein of *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **94**, 322–327 (1997).

Acknowledgements

We thank L. Yamaguchi and Y. Toyoshima for fluorescently labelled axoneme-microtubules, and J. Howard and W. Schief for valuable discussions. This work was supported by the NIH (S.A.E.), Japan Ministry of Education, Science, Sport & Culture (H.H.) and Human Frontier Science Program (S.A.E. and H.H.).

Correspondence and requests for materials should be addressed to S.A.E. (e-mail: endow@duke.edu).

A deeply knotted protein structure and how it might fold

William R. Taylor

Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

The search for knots in protein has uncovered little that would cause Alexander the Great to reach for his sword. Excluding knots formed by post-translational crosslinking, the few proteins considered to be knotted form simple trefoil knots with one end of the chain extending through a loop by only a few residues^{1,2}, ten in the 'best' example³. A knot in an open chain (as distinct from a closed circle) is not rigorously defined and many weak protein knots disappear if the structure is viewed from a different angle. Here I describe a computer algorithm to detect knots in open chains that is not sensitive to viewpoint and that can define the region of the

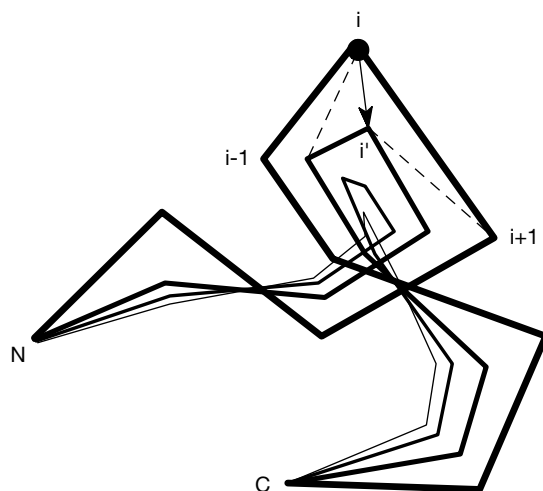


Figure 1 The basic chain smoothing algorithm. Protein chains are drawn as lines connecting the central carbon atom in the backbone of each residue unit running from the amino (N) terminus to the carboxy (C) terminus. Beginning at the second residue, for each residue point (i) in the starting conformation, the average coordinate of $i-1$ and $i+1$ was taken as the new position (i') for the residue. This procedure was then repeated, and the results of this are progressively smoother chains, shown as a series of fainter lines. Note that the termini do not move. With each move, it was checked that the chains did not pass through each other. This was implemented by checking that the triangles $\{i'-1, i, i'\}$ and $\{i, i', i+1\}$ (dashed lines in the Figure) did not intersect any line segment $\{j'-1, j'\}$ ($j < i$) before the move point or any line $\{j, j+1\}$ ($j > i$) following.

chain giving rise to the knot. It characterizes knots in proteins by the number of residues that must be removed from each end to abolish the knot. I applied this algorithm to the protein structure database and discovered a deep, figure-of-eight knot in the plant protein acetohydroxy acid isomeroreductase⁴. I propose a protein folding pathway that may explain how such a knot is formed.

Pulling the ends of a given piece of string will usually decide whether it is knotted or not. Because we hold the ends, the string and our body form a closed circle and there is no danger of untying the knot as it is pulled. The definition of knots in circular strings is mathematically well defined⁵, and one way to approach the problem in open strings is simply to join the ends (as we do when we pick up a string). This is fine for clear knots (where the ends of the string are remote from the knot site), but if the ends are tangled-up together with the knot then any algorithm devised to 'pick-up' the ends creates the risk that the external connections might either untie an existing knot, or create a new one. Fortunately, the ends of protein chains (being charged) tend to lie on the surface of the structure and so can often be joined unambiguously by a wide loop. However, not all termini lie on the surface and if we want to locate the knotted

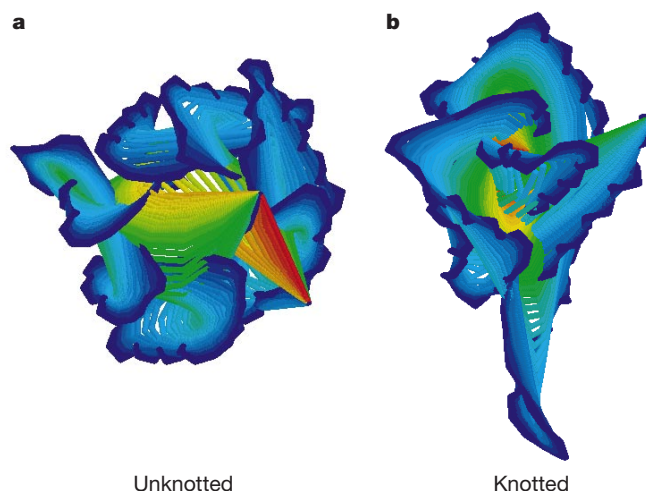


Figure 2 Smoothed protein structures. Applying the smoothing algorithm described in the text (and Fig. 1) to protein structures produces a series of increasingly smoothed chains, coloured from blue to red. (For clarity, the native starting structure is not shown). **a**, Applied to a protein that has no knots (triosephosphate isomerase, (1tph1)) results in a straight line (red) joining the termini. To reach this stage took 52 smoothing iterations. **b**, Applied to the knotted protein (the carboxy-terminal domain of acetohydroxy acid isomeroreductase, (1yvel)), a straight line is never attained and a small knot remains deep in the (red) core part of the protein. This is shown in Fig. 3b. Figure prepared using RasMol version 2.6 (ref. 9).

region, any deletion from the ends could drive the termini deeper.

An alternative approach is to invert the problem: rather than extending the termini outwards, these can be left fixed and the rest of the protein made to shrink around them. This was done by contracting the protein chain as if it were a rubber band: specifically, with residues represented by the coordinates of their α -carbon, each point was repeatedly replaced by the average of itself and its two neighbours (Fig. 1). This procedure, which has been used previously to visualize protein chains, quickly reduces the protein to a smooth curve but leaves the termini untouched (as they do not have two neighbours). If continued indefinitely, all the points will lie on the line connecting the termini. When investigating topological features of the chain, however, it is most important that two parts of the chain cannot pass through each other and this was checked with each move of a residue point. Any move that violated this check was not implemented, leaving the current residue in its original position. With this condition, unknotted strings will still be reduced to a straight line, but those containing knots will become blocked (Fig. 2).

Although this simple algorithm is sufficient to detect knots, some pragmatic refinements were made. Because of a limit in the

Table 1 Knots found in proteins

Protein	PDB code	Length	Knot	Core	Depth
Acetohydroxy acid iso/red	1yvel	513	LRRL	245-444	17,220
Acetohydroxy acid iso/red C-domain	1yvel	234	LRRL	16-215	170
Carbonic anhydrase IV	1zncA	262	RRR	31-261	64
Ubiquitin YUH1-UBAL	1cmxA	214	LLL	5-210	30
VP3 core protein (bluetongue virus)	2btvB	885	LRRL	203-879	1,632
S-adenosylmethionine synthetase	1fugA	383	RRR	10-276	1,070
Carbonic anhydrase	1kopA	223	RRR	39-223	40
Carbonic anhydrase I	1hcb	258	RRR	28-256	87
Carbonic anhydrase V	1dmxA	237	RRR	6-234	28

A selection of 3,440 protein structures was extracted from the current protein structure databank (PDB) in which no two structures shared more than 50% sequence identity. (see ref. 11 for selection protocol). A separate entry is included for the core of the C-terminal domain of 1yvel (excluding the N-terminal domain and the last 50 residues). Residue numbers are the sequential numbering of the PDB file and the size of the protein is tabulated as length. The knots found were characterized by the string formed by the handedness of their successive crossovers (L, left; R, right). Note that while the two trefoils are distinct even as circular knots, the others can be created by introducing different break-points in a circular knot. The core of the knot was determined by a series of deletions from each terminus to find the smallest region that remained knotted under application of the method, and the product of the number of residues that must be deleted from the ends to free the knot is tabulated under depth. Note that the knot type in the minimal core of 1yvel is not the same as in the full C-terminal domain (or full protein), as the deletion of most of the first α -helix in the domain results in the chain being 'pulled' tight from a different direction. The knot in these more complete structures is the figure-of-eight type (LRRL) depicted in Fig. 3b.

numerical accuracy of the computer calculations, each line between residues was represented by an impenetrable tube (1 Å diameter). This allowed residues to be progressively removed: specifically, the middle residue of three that were almost collinear or where the separation of the outer two fell within the tube diameter. This not only improved execution time but led to an even simpler test for knots as any chain that can be reduced to just its two termini is not knotted. Those with more than two residues remaining are either knots or tangles in which a group of moves have become 'grid-locked' (like rush-hour traffic at junction). This latter condition was eased (but not completely eliminated) by making a slight reduction in the tube diameter any time the chain became stuck.

In practice, most chains of a few hundred residues are reduced to their termini in around 50 iterations. If by 500 iterations a chain was still not reduced two points, then the resulting configuration was analysed in more detail. As the termini are now well separated from the knot-site, they can be unambiguously joined and analysed as a 'proper' circular knot. This can be done using one of the knot-invariant polynomials, such as the Alexander or Jones polynomials^{1,5}. However, the few knots encountered in proteins are so simple that they do not require any sophisticated analysis. Furthermore, from a theoretical perspective, not only are protein knots directional but also they have a unique break-point (between the termini) which is not taken into consideration by any of the polynomial forms. As a working tool, a simpler method was adopted to characterize these open knots based on the Dowker knot notation⁵ (see legend to Table 1).

Applying this method to a non-redundant selection of protein structures (see Table 1 for selection details) revealed a surprising number of knots. A few of these proved to be unresolved tangles (often forming slip-knots), and some others were caused by breaks in the chain creating an unnatural short-cut. The former were all eliminated by running the program with a smaller 'tube' diameter but the latter could only be removed through visual inspection. Of the seven remaining structures (Table 1), five were right-handed trefoils including related carbonic anhydrase

structures (1zncA, 1kopA, 1hcb, 1dmxA) and the protein S-adenosylmethionine synthetase (1fugA) both of which had been identified previously. In addition, three new knots were found: a left-handed trefoil in ubiquitin (1cmxA), and two figure-of-eight knots in a viral core protein (2btvB) and acetohydroxy acid isomeroreductase (1yveI; Fig. 3b). These last two are of particular interest as they include an additional crossover above the trefoil and are therefore less likely to be formed by a wandering chain during folding. This was confirmed by simulation of random and semi-random compact protein-like chains in which the trefoil was by far the most common knot type (data not shown). The location of the two figure-of-eight knots was determined by a series of deletions from both termini of the protein chain. This revealed that the knot in 2btvB required just the last eight residues, which is similar to the deepest trefoil knot. By contrast, the knot in 1yveI, which is contained in the carboxy-terminal domain of the protein, remained until 70 residues were deleted from the carboxy terminus and 245 residues (including a complete domain) were removed from the amino terminus (Fig. 3a).

It is interesting to speculate how a protein with such a deep and complex knot might fold—as it is difficult to imagine over 50 residues being 'fed' through a loop in a reproducible way during folding. Clues to the folding of this protein can be found in a clear internal duplication within the domain, comprising 88 residue pairs with 2.0 r.m.s. deviation (as measured by the program SAP (ref. 6) over the α -carbon positions). If it is assumed that the two most deeply buried, symmetrically equivalent helices initially pack together (A1 and B1, Fig. 4), then the remaining parts of each repeat (A2, A3 and B2, B3) can wrap around this core requiring only that the C-terminal segment can pass through the large loop between the repeats before this contacts and finally packs onto the core (Fig. 4). The symmetry in this arrangement suggests that the protein might have evolved from an exchange of structure or 'swap'⁷ between the two duplicated domains in which the first

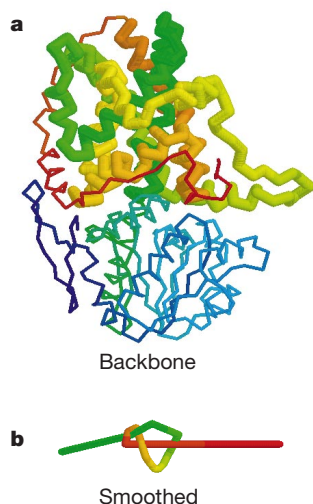


Figure 3 The knot in 1yveI. **a**, Backbone representation of the complete native protein structure (coloured from blue to red in the direction of the chain) with the core of the knotted domain drawn thickened. This region is preceded by a complete nucleotide binding domain (blue-green) and followed by a long loop (red) that wraps around the domain. **b**, The knotted core in the smoothed representation of 1yveI (Fig. 2b), coloured as in **a**. The figure-of-eight knot can be seen clearly. This form was attained after 50 cycles, and if continued an irreducible core consisting of eight points was attained. Figure prepared using RasMol version 2.6 (ref. 9).

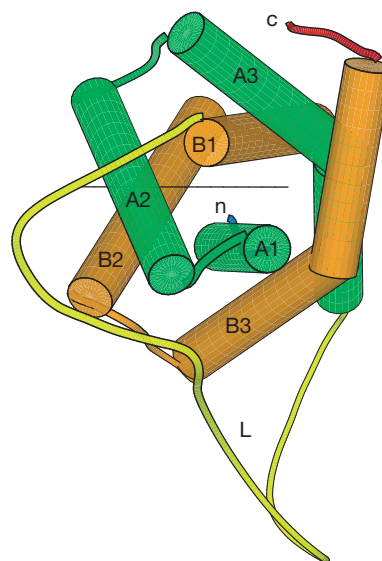


Figure 4 The core of the C-terminal domain in 1yveI. The duplicated portions of the chain (A1–A3 and B2–B3) are viewed down the packing axis of the two central α -helices. (The fine horizontal line between these is a least-squares fit to the pseudo-two-fold axis. The following parts of the duplicated regions (A2, A3 and B2, B3) wrap around this central pair creating the core of the knot. The final tying of the knot requires that the C-terminal part of the chain (including 45 residues following the point labelled c) passes through the long loop connecting the duplications (labelled L). The origin of this structure can be imagined as a gene duplication giving rise to a linked dimer in which the two central helices then exchanged positions. Figure prepared using MOLSCRIPT version 2.1.2 (ref. 10).

helix in the repeat has been transposed across the twofold axis of symmetry so creating the knot (see ref. 8 for a wider review of related processes). Intriguingly, the best example of a trefoil knot (in IfugA) appears to have arisen in a similar manner, in which a β -strand on the edge of a sheet has been transferred from one duplicated domain to another. Although it cannot be stated that significant knots in proteins will not arise by other means, it appears that the swapping of elements of secondary structure between duplicated domains can provide a source of knotted proteins. $\square$

Received 3 April; accepted 15 June 2000.

1. Mansfield, M. L. Are there knots in proteins? *Nature Struct. Biol.* **1**, 213–214 (1994).
2. Mansfield, M. L. Fit to be tied. *Nature Struct. Biol.* **4**, 116–“7 (1997).
3. Takusagawa, F. & Kamitori, K. A real knot in protein. *J. Am. Chem. Soc.* **118**, 8945–8946 (1996).
4. Biou, V. *et al.* The crystal structure of plant acetohydroxy acid isomeroreductase complexed with NADPH, two magnesium ions and a herbicidal transition state analog at 1.65 Å resolution. *EMBO J.* **16**, 3405–3415 (1997).

5. Adams, C. C. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots* (Freeman, New York, 1994).
6. Taylor, W. R. Protein structure alignment using iterated double dynamic programming. *Protein Sci.* **8**, 654–665 (1999).
7. Bennet, M. J., Schlunegger, M. P. & Eisenberg, D. 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455–2468 (1995).
8. Heringa, J. & Taylor, W. R. Three-dimensional domain duplication, swapping and stealing. *Curr. Opin. Struct. Biol.* **7**, 416–421 (1995).
9. Sayle, R. & Milner-White, E. J. RasMol: Biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374 (1995).
10. Kraulis, P. J. MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
11. Taylor, W. R. Multiple sequence threading: an analysis of alignment quality and stability. *J. Mol. Biol.* **269**, 902–943 (1997).

Correspondence and requests for materials should be addressed to W.R.T. (e-mail: wtaylor@nimr.mrc.ac.uk). The program used in this work can be found at <ftp://mathbio.nimr.mrc.ac.uk>.